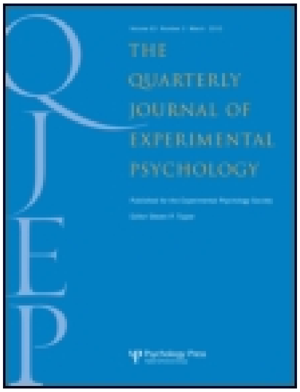


The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning

Lauren L. Emberson [a] [b] , Christopher M. Conway [c] & Morten H. Christiansen [a]

[a] Department of Psychology, Cornell University, Ithaca, NY, USA

[b] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA

[c] Department of Psychology, Saint Louis University, St. Louis, MO, USA

Available online: 22 Feb 2011

To cite this article: Lauren L. Emberson, Christopher M. Conway & Morten H. Christiansen (2011): Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning, The Quarterly Journal of Experimental Psychology, 64:5, 1021-1040

To link to this article: http://dx.doi.org/10.1080/17470218.2010.538972

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2011, 64 (5), 1021–1040

Psychology Press
Taylor & Francis Group

Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning

Lauren L. Emberson[1,2], Christopher M. Conway[3], and Morten H. Christiansen[1]

[1]Department of Psychology, Cornell University, Ithaca, NY, USA
[2]Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA
[3]Department of Psychology, Saint Louis University, St. Louis, MO, USA

Implicit statistical learning (ISL) is exclusive to neither a particular sensory modality nor a single domain of processing. Even so, differences in perceptual processing may substantially affect learning across modalities. In three experiments, statistically equivalent auditory and visual familiarizations were presented under different timing conditions that either facilitated or disrupted temporal processing (fast or slow presentation rates). We find an interaction of rate and modality of presentation: At fast rates, auditory ISL was superior to visual. However, at slow presentation rates, the opposite pattern of results was found: Visual ISL was superior to auditory. Thus, we find that changes to presentation rate differentially affect ISL across sensory modalities. Additional experiments confirmed that this modality-specific effect was not due to cross-modal interference or attentional manipulations. These findings suggest that ISL is rooted in modality-specific, perceptually based processes.

Keywords: Implicit learning; Statistical learning; Temporal processing; Multisensory processing; Perceptual grouping.

Implicit statistical learning (ISL) is a phenomenon where infant and adult behaviour is affected by complex environmental regularities seemingly independent of conscious knowledge of the patterns or intention to learn (Perruchet & Pacton, 2006). Because young infants are sensitive to statistical regularities, ISL has been argued to play an important role in the development of key skills such as visual object processing (Kirkham, Slemmer, & Johnson, 2002) and language learning (Saffran, Aslin, & Newport, 1996; Smith & Yu, 2008). Underscoring its importance for development and skill acquisition, ISL has been observed using a wide range of stimuli from different sensory modalities and domains (nonlinguistic auditory stimuli: Saffran, 2002; Saffran, Johnson, Aslin, & Newport, 1999; tactile stimuli: Conway & Christiansen, 2005; abstract visual stimuli: Fiser & Aslin, 2001; Kirkham et al., 2002). Together, these findings indicate that ISL is a domain-general learning ability spanning sense modality and developmental time.

Given that ISL occurs with perceptually diverse input, many influential models and theories of ISL

Correspondence should be addressed to Lauren L. Emberson, Department of Psychology, 211 Uris Hall, Cornell University, Ithaca, NY, USA. E-mail: lle7@cornell.edu

http://www.psypress.com/qjep
DOI:10.1080/17470218.2010.538972

have presupposed a mechanism that treats all types of input stimuli (e.g., tones, shapes, syllables) as equivalent beyond the statistical structure of the input itself (e.g., Altmann, Dienes, & Goode, 1995; Perruchet & Pacton, 2006; Reber, 1989; Shanks, Johnstone, & Staggs, 1997). While great strides have been made under this equivalence assumption, there is evidence, contrary to this view, that ISL is *not* neutral to input modality. Instead, the perceptual nature of the patterns appears to selectively modulate ISL.

In this paper, we employ a known perceptual phenomenon to examine ISL under different perceptual conditions. Specifically, we manipulated the temporal distance of successive stimuli in auditory and visual ISL streams. The perceptual literature predicts that changes of temporal distance will have opposite effects on auditory and visual processing. If ISL were also differentially affected by temporal distance, this would suggest that the mechanisms mediating ISL do not in fact treat all types of perceptual input equivalently.

In addition, we investigated the role of selective attention in modifying learning under these different perceptual conditions. While previous research has suggested that selective attention can compensate for perceptual effects in ISL (e.g., Baker, Olson, & Behrmann, 2004; Pacton & Perruchet, 2008), this claim has only been tested in a small range of perceptual conditions in the visual modality only. Here we examine whether selective attention can compensate for large differences in rate of presentation in both the visual and the auditory modality. Specifically, we predict that while selective attention may be able to support learning amidst mild disruptions to perceptual processing (as in Baker et al., 2004), attention is not sufficient to overcome more substantial changes in perceptual conditions like those explored in the current study.

In sum, we manipulated attention to auditory and visual streams under temporally proximal and distal conditions in order to examine what effect changes of presentation rates have on auditory and visual ISL. If the mechanisms of ISL are sensitive to the perceptual nature of stimulus input beyond statistical structure, then we predict that rate and modality will interact to affect learning outcomes.

Modality effects in implicit statistical learning

While ISL is perceptually ubiquitous, with adults and infants able to detect statistical regularities in multiple sensory modalities, recent studies with adult learners have pointed to systematic differences in ISL across these modalities (Conway & Christiansen, 2005, 2006, 2009; Robinson & Sloutsky, 2007; Saffran, 2001). Specifically, modality differences in ISL appear to follow the visual:spatial::auditory:temporal characterization seen in other perceptual and cognitive tasks, where spatial and temporal relations are processed preferentially by the senses of vision and audition, respectively (Kubovy, 1988).

While temporal and spatial information are both important for visual and auditory processing, these sources of information appear to play different roles across perceptual systems. The visual:spatial::auditory:temporal analogy (Kubovy, 1988), used to explain auditory and visual processing differences, has its roots in the nature of sensory objects. Sound is a temporally variable signal, and, since sounds do not persist, their locations in space are ephemeral. Conversely, visual objects are more spatially constant. Thus, it is adaptive for auditory processing to be more sensitive to the temporal aspects of environmental information (Chen, Repp, & Patel, 2002) whereas the adult visual system appears to preferentially encode spatial information (Mahar, Mackenzie, & McNicol, 1994). Furthermore, the visual:spatial::auditory:temporal characterization extends beyond perceptual tasks to memory (serial recall: Penney, 1989).[1]

[1] The range of visual processing explored in the current paper is restricted: We are examining visual processing and learning of sequentially presented, unfamiliar abstract shapes. Other visual tasks have revealed the visual system to have sophisticated temporal processing (e.g., rapid serial visual presentation of scenes and photographs in Potter, 1976). However, with the current visual task, it is well established that visual processing is relatively poor especially when compared to auditory processing.

These differences in processing between auditory and visual systems are also present in ISL. Consistent with a spatial bias in visual processing, visual learning is facilitated when stimuli are arrayed spatially (Conway & Christiansen, 2009; Saffran, 2002). When stimuli are presented in a temporal stream, auditory learning is superior to vision (Conway & Christiansen, 2005). These findings point to important differences in the ways in which auditory and visual statistical patterns are learned.

We propose that comparisons of learning across perceptual modalities help elucidate the nature of the mechanism(s) underlying ISL. Moreover, these modality effects in ISL may indicate that the underlying mechanisms are sensitive to the perceptual nature of the input beyond statistical structure. One could think of these mechanisms as being "embodied" (Barsalou, Simmons, Barbey, & Wilson, 2003; Conway & Christiansen, 2005; Glenberg, 1997) where the learning mechanisms are situated in the perceptual process itself.

Modality-specific perceptual grouping and ISL

Modality differences can also be conceptualized through the lens of Gestalt perceptual-grouping principles. The spatial bias in visual processing has been formalized by the "law of proximity": Visual stimuli occurring close together in space are perceptually grouped together as a single unit (Kubovy, Holcombe, & Wagemans, 1998; Wertheimer 1923/1938), with the strongest grouping occurring in spatially contiguous visual objects (Palmer & Rock, 1994). Analogously, sounds that are presented closer together in time are more likely to form a single perceptual unit or stream (Handel, Weaver, & Lawson, 1983). A logical consequence of the law of proximity is that sounds that are far apart in time, and visual stimuli that are far apart in space, will fail to form perceptual units (Bregman, 1990). For example, previous research has indicated that sounds presented more than 1.8–2 s apart are not perceived as part of the same stream of sounds (Mates, Radil, Müller, & Pöppel, 1994) and that the visual system fails to group objects together as the space between them increases (Palmer & Rock, 1994).

Recently, Baker et al. (2004) examined the impact of spatial perceptual grouping on visual ISL. Participants were presented with statistical patterns of simultaneously presented pairs of visual shapes; pairs were either spatially connected by a bar (a strong form of visual perceptual grouping) or not. They found that participants in the stronger perceptual grouping condition had better learning than those in the weaker perceptual grouping conditions. Similar results have been found by Pacton and Perruchet (2008). These studies demonstrate that spatial perceptual grouping conditions affect visual ISL.

To date, the relationship between perceptual grouping and learning in the auditory modality has not been systematically investigated. If strong perceptual grouping aids ISL, then auditory perceptual grouping ought to improve as sounds are presented at closer temporal proximity (i.e., at a faster rate). Conway and Christiansen (2009) reported that increasing rates of presentation from 4 stimuli/second (250-ms stimulus onset asynchrony, SOA) to 8 stimuli/second (125-ms SOA) did not impact learning in the auditory modality. However, this is a small range of presentation rates, with both rates being well within the limits of auditory perceptual grouping (i.e., SOA less than 2 s). In order to more directly assess the effects of temporal perceptual grouping, more varied grouping conditions need to be examined for both auditory and visual input.

Current experiments

The current paper examines the effect of perceptual grouping along the temporal dimension using greater changes in presentation rate than have been previously investigated. Specifically, the current experiment examines both visual and auditory ISL when the streams are presented either at fast rates of presentation (similar to rates used in previous studies) or under much slower rates of presentation. If auditory ISL is aided by temporal perceptual grouping, auditory learning should improve when sounds are presented closer together in time (i.e., at a faster rate) and should be disrupted when sounds are presented further apart in time (i.e., at a slower rate).

In contrast, we predict the opposite effect of presentation rate on visual ISL: Since visual processing has poorer temporal resolution, visual ISL should not be facilitated by a fast rate of presentation as auditory ISL would. Instead, visual ISL will improve with slower rates of presentation because this is less temporally demanding on the visual system. Previous work has demonstrated improvements to visual ISL with slower rates of presentation (Conway & Christiansen, 2009; Turk-Browne, Jungé, & Scholl, 2005).

It is crucial to note that the changes in temporal rate employed in the current study do not obfuscate the individual stimuli themselves. At the fastest rate of presentation employed in the current study, previous work (Conway & Christiansen, 2005) as well as pilot testing revealed that there is robust perception of individual visual and auditory stimuli. Thus, by “changes in perceptual conditions” we are not referring to changing the ability of participants to perceive individual stimuli. However, as reviewed above, changes in rate of presentation have been shown to affect perception of auditory stimuli as occurring in a single stream and to decrease ability of the visual system to resolve streams of stimuli. Thus, it is the perception of these streams of stimuli, in which statistical regularities are presented, but not the individual stimuli that is being affected by differences in rate of presentation.

In the current paradigm, participants are familiarized with both visual and auditory statistical regularities. Conway and Christiansen (2006) observed that statistical information from two different streams could be learned simultaneously if these streams were from different modalities (visual and auditory) but not if they were instantiated in perceptually similar stimuli. In their design, strings of stimuli were generated by two different artificial grammars and interleaved with one another, as complete strings, in random order. In the current study, we investigated statistical learning of triplets of stimuli within a single stream (Figure 1a). Since triplet boundaries are key statistical information, alternating between full triplets would provide an explicit boundary cue. To avoid such a scenario while presenting both auditory and visual triplets, we adapted the interleaved design from Turk-Browne et al. (2005) to present an auditory and a visual familiarization stream (see Figure 1b for illustration of the interleaved design as applied to the current study). In addition, interleaving two familiarization streams avoids cross-modal effects in ISL that have been observed when visual and auditory streams are presented simultaneously (Robinson & Sloutsky, 2007).

Thus, if ISL is affected by modality-specific or perceptual processes, we predict that rate manipulations will have opposite effects on visual and auditory ISL: (a) We expect auditory ISL to be poorer at slower rates of presentation than learning at fast rates, and (b) we predict the opposite pattern of results in the visual modality: We expect learning to be stronger when presentation rates are slow than learning of visual elements presented at fast presentation rates.

In addition to manipulating the rate of presentation in the current study, we also manipulate selective attention to the streams. While the necessity of attention is unclear in ISL (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), it has recently been established that selective attention to the information containing the statistical regularities boosts performance in both the visual and the auditory modalities (Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne et al., 2005). Consistent with this work, we predict that there will be significantly reduced learning for the unattended streams for both visual and auditory sensory modalities with both rates of presentation. Thus, we do not expect to see an effect of rate in the unattended streams given that we anticipate seeing no learning in conditions without attention.

Focusing on predictions for the attended streams, it has been proposed that one way in which attention aids in ISL is through boosting performance when perceptual grouping conditions are unfavourable. Recent work has suggested that poor perceptual grouping conditions can be

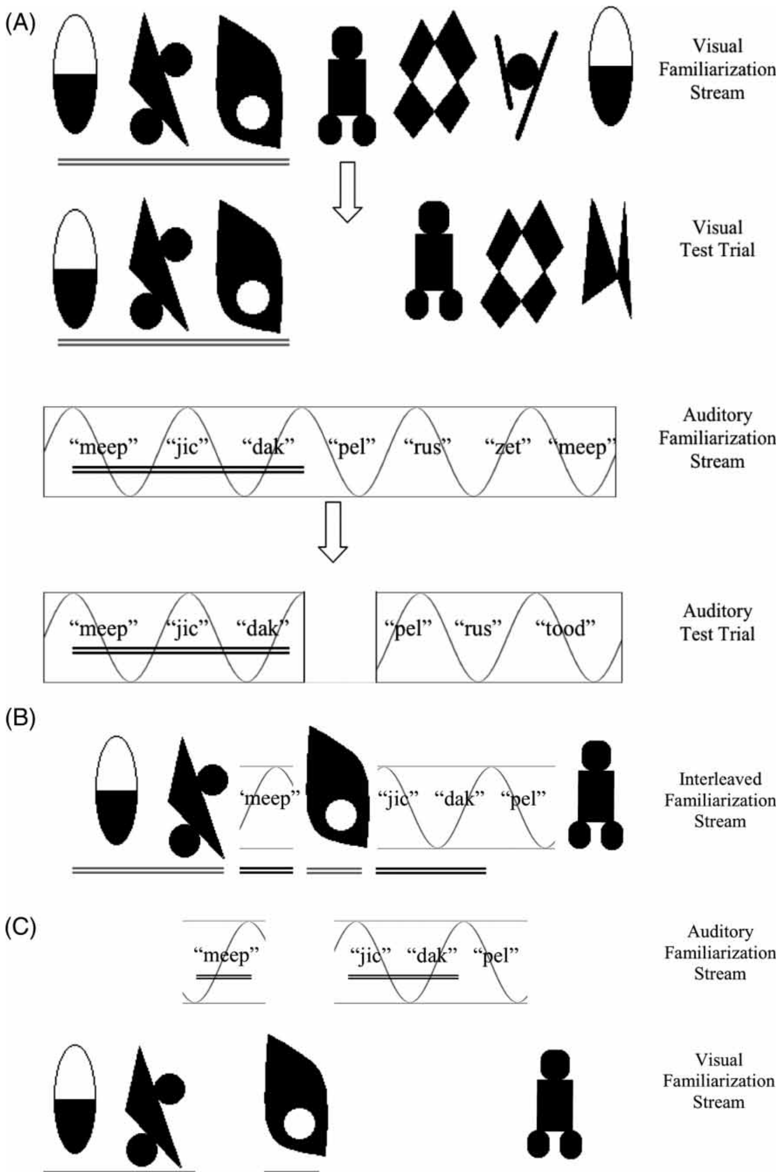


Figure 1. *(A) A sample of separate visual and auditory familiarization streams prior to interleaving. A sample triplet is underlined in each stream (visual: grey; auditory: black). Test trials compared a triplet and foil from a single modality. (B) In Experiments 1 and 2, visual and auditory streams were interleaved so stimuli from both modalities were presented sequentially with presentation pseudorandomly switching between streams with no more than six consecutive elements from a single modality. (C) In Experiment 3, interleaved streams were presented with the same timing of presentation for a stream from an attended modality but with unattended stimuli from the other modality removed.*

overcome with selective attention to relevant stimuli (Baker et al.; 2004; Pacton & Perruchet, 2008). However, the type and range of perceptual grouping in these studies has been limited, and investigations have not extended beyond the visual modality. It is unknown whether selective attention can overcome poor grouping conditions in the auditory modality and whether attention is always sufficient to overcome even extreme disruptions in perceptual grouping.

Given the large variations in temporal rate in the current studies, we predict that selective attention will not be sufficient to compensate for the poor perceptual conditions induced by these changes in presentation rate. Thus, we expect to see that the modality-specific effect of temporal rate (i.e., poor at fast rates for visual and poor at slow rates for auditory) will persist even if participants selectively attend to these modalities. An interaction of rate and modality under conditions of selective attention would be evidence that selective attention is not always sufficient to compensate for poor perceptual conditions.

EXPERIMENT 1: INTERLEAVED, FAST PRESENTATION (375-ms SOA)

To examine the modality-specific effects of temporal perceptual grouping (rate of presentation), we interleaved two familiarization streams governed by statistical information in the visual and auditory modalities. The current experiment presented streams at a rate similar to that in previous ISL studies (SOA less than 500 ms). As with this previous work, we predict an auditory superiority effect in ISL at these relatively fast rates of presentation (Conway & Christiansen, 2005, 2009; Saffran, 2002).

Two familiarization streams (auditory and visual) were interleaved to create a single stream; this was done by sampling one to six elements at a time from a single stream consecutively (see Figure 1b). Interleaving streams resulted in a predictable set of transitional probabilities that was roughly equal across experimental groups (Table 1). Transitional probabilities are higher for successive elements within triplets than for those spanning triplets, providing a cue for learning (e.g., see Fiser & Aslin, 2001; Saffran et al., 1996; Turk-Browne et al., 2005).

As with Turk-Browne et al. (2005), selective attention was manipulated between streams. While some research has indicated that explicit attention to stimuli is not required for ISL (Saffran et al., 1997), other research has demonstrated that selective attention *aids* in ISL in both the visual (Turk-Browne et al., 2005) and the auditory (Toro et al., 2005) modalities. Thus, we do not expect to see evidence of learning in unattended streams regardless of rate of presentation.

Method

Participants

Thirty-two participants were recruited from psychology classes at Cornell University, earning extra credit or $10/hour. All participants reported normal or corrected-to-normal vision and no serious auditory deficits or neurological problems.

Materials

Auditory and visual stimuli were presented at a rate similar to that in previous statistical learning studies (e.g., Conway & Christiansen, 2005; Saffran et al., 1996, 1997): Visual and auditory

Table 1. *Transitional probabilities of elements in the stream for each modality in isolation and interleaved*

	Isolation	*Interleaved*	*Isolation*	*Interleaved*
p(any particular shape), e.g., p(B)	1/5 × 1/3	1/15 × 1/2	.064	.032
p(any repeated shape), e.g., p(A)	1/5 × 1/3	1/15 × 1/2	.068	.034
p(any pair within a triplet), e.g., p(A, B)	1/15 × 1/1	1/30 × 1/2 × 1/1	.064	.016
p(any pair spanning triplets), e.g., p(C, G)	1/15 × 1/4	1/30 × 1/2 × 1/4	.016	.004
p(any given triplet), e.g., p(A, B, C)	1/15 × 1/1 × 1/1	1/30 × 1/2 × 1/1 × 1/2 × 1/1	.064	.008
p(any given nontriplet), e.g., p(B, C, G)	1/15 × 1/1 × 1/4	1/30 × 1/2 × 1/1 × 1/2 × 1/4	.016	.004
p(any foil sequence), e.g., (A, B, I)	0	0	0	0

Note: As observed by participants in Experiments 1 and 2. Elements: monosyllabic nonwords or shapes. Modalities: auditory or visual, respectively.

stimuli are presented for 225 ms with an interstimulus interval (ISI) of 150 ms, resulting in an SOA of 375 ms. All stimuli were presented using E-prime stimuli presentation software (Version 1, Psychology Software Tools).

Visual stimuli. Fifteen novel abstract shapes were drawn using MS Paint for Windows 98 Second Edition (see Appendix A). The stimuli were designed to be perceptually distinct and not easily labelled verbally. During central presentation, shapes measured 4 cm by 6 cm on a 17-inch Samsung SyncMaster 955DF. Participants were seated 65 cm from the screen.

Auditory stimuli. Fifteen monosyllabic nonwords, recorded by a female, native English speaker, were chosen to obey the phonological rules of English and be easily distinguishable from each other but as unique and unfamiliar as possible (see Appendix B). All nonwords were edited using Audacity for OSX (Version 1.2.2, Free Software Foundation, Boston, MA; Audacity Team, 2005).

Procedure

Participants were randomly assigned to one of three groups: two experimental groups, *visual attention* or *auditory attention* (24 participants), or *nonfamiliarized controls*. Participants in the two experimental groups had identical procedures except for the inclusion in the instructions that participants preferentially attend to a single modality.[2] Immediately following familiarization, experimental participants were tested for evidence of learning in both the visual and the auditory modalities. Participants in the *nonfamiliarized control* group were given the same testing procedure as were those in the experimental condition without receiving familiarization.

Familiarization. Stimuli were grouped offline into single-modality triplets resulting in five auditory and five visual triplets. In order to ameliorate any effects of triplet grouping, multiple groupings were used across participants with each triplet grouping employed in all conditions. Thirty presentations of each triplet were randomly ordered such that no triplet or pairs of triplets were immediately repeated (e.g., ABCABC or ABCDEFABCDEF). A cover task was employed: Participants were asked to detect repeated elements in the familiarization stream using a button box, and no feedback was given. The first and third elements of each triplet were repeated two times during familiarization (e.g., ABCCDEFGGHI; Turk-Browne et al., 2005).

Auditory and visual familiarization streams were pseudorandomly interleaved by sampling each stream in order and without replacement with no more than 6 elements from one stream sampled consecutively (see Figure 1b). Critically, the process of interleaving did not highlight the triplet structure of the familiarization streams, with streams often switching between modalities within triplets. This resulted in a familiarization stream of 940 elements: 470 from each modality. Participants were given a self-timed break halfway through familiarization. The sequence of interleaving was counterbalanced such that the interleaved order of the visual elements for one group of participants was that of the auditory elements for another group of participants; attention was counterbalanced across modality and interleaved order.

Testing. Test trials were constructed for each modality separately comparing triplets from familiarization to foils (Figure 1a). Then test trials from both visual and auditory test trials were presented in random order in a multimodal testing block. Within each modality, the testing phase

[2] Before familiarization, participants were instructed to attend to a single modality (auditory or visual) depending on their assigned group. They were instructed that stimuli in the other modality were meant to provide distraction. Participants were told to respond to the repeated elements in their assigned modality only. If participants were in the auditory attention group, they were specifically instructed to still look at the monitor but to just direct their attention to the auditory stimuli. Due to a data collection error, repeat responses were not collected. However, the replication of these results in Experiment 3 without unattended stimuli indicates (a) that participants are in fact attending to the assigned sensory modality and (b) that attention to a particular modality was analogous to attention during exposure without unattended stimuli (i.e., there was no interference).

consisted of a forced-choice task pairing the five triplets constructed for each participant with five foils and counterbalanced for order of presentation, resulting in 50 test trials per modality (5 triplets × 5 foils × 2 order). The same foils were paired with all triplets during test; thus there were the same number of foils and triplets used at test to equate exposure. Foils were constructed from the same shapes and nonwords, designed to violate the triplet structure but not absolute element position (e.g., triplet: ABC, DEF, GHI; foil: ABF, DEI, GHC). All of these stimuli were presented in the same manner and with the same timing as the familiarization stream. Foils and triplets were separated by 1,000 ms of silence. Following the methodology of Conway and Christiansen (2005) and Saffran (2001), participants were instructed to report which triplet seemed "more *familiar* or *right* based on [their] previous task, if applicable". They were instructed to respond to the triplet and not the individual elements. After presentation of a pair of test items, participants were prompted to press Key 1 (of a 4-key response pad) if they felt that the first item was more "familiar" or "right" and to press Key 4 for the second item. The response screen was self-timed and participants received no feedback on their responses. Participants were instructed that there was no order to the modality of successive test trials. The dependent measure was accuracy in discriminating triplets from foils across 50 test trials.

Results

Results are collapsed across both interleaved pattern and triplet groupings with analysis occurring only along dimensions of experimental groups (auditory vs. visual attention) and experimental versus nonfamiliarized controls.

Nonfamiliarized controls

Performance of participants in the control group was evaluated against chance performance (25 out of 50, or 50%). Control participants performed at 49% accuracy for both modalities, and neither was significantly different from chance performance: visual, $t(7) = -0.36$, $p = .73$; auditory, $t(7) = -0.80$, $p = .45$.

Experimental groups

Participants who attended to auditory stimuli correctly responded to 63% of auditory test trials and 54% of visual test trials. Those who attended to visual stimuli correctly responded to 57% of visual test trials and 47% of auditory test trials (see Figure 2). Comparing experimental performance to control, only the attended auditory condition differed significantly from nonfamiliarized controls, $t(18) = 5.95$, $p < .001$; auditory unattended, $t(18) = -0.420$, $p > .5$; visual attended: $t(18) = 1.73$, $p = .10$; visual unattended: $t(18) = 1.336$, $p = .20$.

Effects of attention. To specifically investigate the effects of selective attention in the interleaved-multimodal design, planned t tests were performed to compare performance for a single modality in attended and unattended conditions, across experimental groups. This comparison of attended and unattended streams yielded a significant difference in the auditory modality only: auditory attended versus unattended, $t(22) = 4.16$, $p < .01$; visual attended versus unattended, $t(22) = 0.90$, $p = .38$.

Modality effects. Experimental data were submitted to a two-way analysis of variance (ANOVA; visual vs. auditory attention, within-subject factor: visual vs. auditory presentation). While there is no main effect of modality, $F(1, 22) = 0.056$, $p > .5$, there

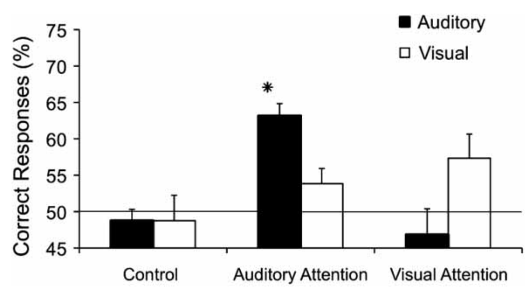


Figure 2. *Mean test performance (percentage correct out of 50) from Experiment 1. Visual and auditory ISL (implicit statistical learning) performance is presented for control, unattended, and attended conditions at fast presentation rate (375-ms stimulus onset asynchrony, SOA).*

is a significant modality by attention interaction, $F(1, 22) = 16.21$, $p = .001$. That is, modality effects were obtained specifically when participants were devoting attention to a given input stream. While direct tests of attended performance across modalities do not reveal a significant difference, $t(22) = 1.573$, $p > .1$, the interaction of modality and attention indicates that modality of presentation is not uniformly affecting learning across attentional conditions. Together with the results presented earlier, a significant effect of attention in the auditory modality only and significant learning is restricted to the attended auditory stream, these results indicate that auditory ISL is superior to visual ISL at this rate of presentation when selective attention is deployed. Increased ISL in the auditory modality is consistent with previous findings using similarly timed rates of presentation (e.g., Conway & Christiansen, 2005).

Discussion

Here we used a multimodal interleaved design to investigate auditory and visual ISL. This experimental design is a novel combination and extension of that used by Conway and Christiansen (2006) and Turk-Browne et al. (2005). Our results corroborate previous cross-modal ISL findings. First, using similar rates of presentation in the current study, auditory ISL appears to have superior performance to visual ISL (Conway & Christiansen, 2005; Robinson & Sloutsky, 2007; Saffran, 2002). Second, concerning the effect of attention, our results are again consistent with previous studies showing that attention can improve learning (Toro et al., 2005; Turk-Browne et al., 2005). However, a significant interaction was obtained, indicating that selective attention improved auditory learning more than visual learning, which remained at control-level performance whether or not selective attention was deployed. Thus, at this relatively fast presentation rate, only auditory learning occurred, even when selective attention was available. Under the same presentation conditions, we do not find evidence of visual learning even with the aid of selective attention. This is likely because, while individual stimuli are easily perceived at the current rate of presentation, visual processing has relatively poor temporal resolution in the current task. See the introduction for a more in-depth discussion.

EXPERIMENT 2: INTERLEAVED, SLOW PRESENTATION (750-ms SOA)

The results from Experiment 1 are consistent with those from previous studies demonstrating superior auditory learning at fast presentation rates (when the input is attended). In the current experiment, we move beyond the temporal distances previously explored in the ISL literature by increasing the distance between successive elements from 375-ms SOA to 750-ms SOA, effectively increasing the amount of time between successive elements in the presentation stream. In fact, given the interleaved design and the increased rate of presentation, the average amount of time between successive visual-to-visual or auditory-to-auditory elements is 2.25 s.[3] Thus, this rate of presentation provides input conditions that are beyond the perceptual grouping tolerance of the auditory system (Mates et al., 1994). See Figure 3 for an illustration of the relative length of pauses for a single element (average is 3 elements) in Experiment 1 (top panel) and Experiment 2 (centre panel) relative to the length of pause necessary to produce significant temporal grouping disruption (bottom panel).

Based on our previous discussion, this slower rate should have opposite effects on visual and auditory ISL. Given that weak spatial perceptual grouping can reduce visual ISL (Baker et al., 2004), we predict a similarly negative effect for weak temporal

[3] In the current experimental methods, there were between 1 and 6 stimuli from a single familiarization stream presented consecutively. The mean number of consecutive stimuli was 3, which, at the rate of presentation employed in Experiment 2, has a duration of 2.25 s. Thus, the average length of pause in an attended familiarization stream, caused by presentation of the unattended familiarization stream, was 2.25 s.

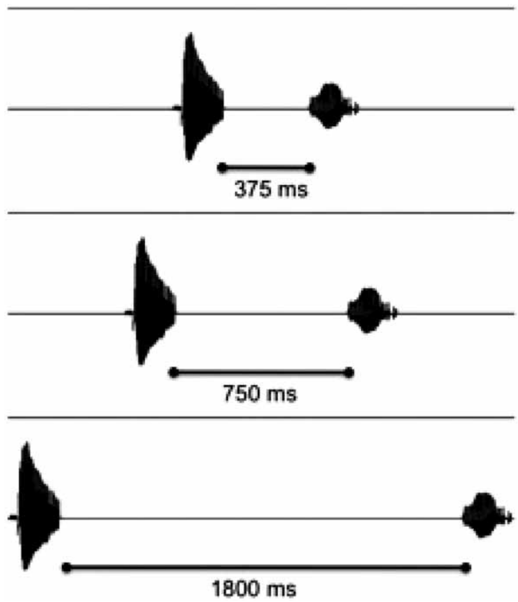


Figure 3. *Illustration of the temporal separation created by the interleaving of a single unattended element at the fast (375-ms stimulus onset asynchrony, SOA) and slow (750-ms SOA) presentation speeds in relation to the limits of auditory temporal perceptual grouping (1.8–2 s). On average, 3 unattended elements were presented consecutively.*

perceptual grouping on auditory ISL. Thus, we predict that a slow rate of presentation will have a negative effect on auditory ISL. However, given that the visual system has relatively poor temporal processing in the current task (see Footnote 1), a decreased rate of presentation should have a positive effect on visual ISL because it places less of a demand on the visual system than does the fast presentation rate used in Experiment 1.

Method

All methods, materials, and procedures were identical to those in Experiment 1 with the exception of presentation rate: Both visual and auditory stimuli were present for 450 ms with a 300-ms ISI (750-ms SOA). To accommodate a slower rate while maintaining natural production, a largely overlapping set of monosyllabic nonwords were recorded by a female, native-English speaker (see Appendix B) and were edited to 750-ms SOA.

Another 32 participants were randomly assigned to one of three groups: two experimental groups (*visual attention* or *auditory attention*), or *nonfamiliarized controls*.

Results

Nonfamiliarized controls responded correctly to 43% of the visual test trials and 46% of the auditory test trials; neither result was significantly different from control performance at 50%: visual, $t(7) = -1.27$, $p = .25$; auditory, $t(7) = -1.42$, $p = .20$ (Figure 4). Participants who attended to the visual modality correctly responded to 65% of visual test trials and 48% of auditory test trials. Those who attended to the auditory modality correctly responded to 55% of auditory test trials and 52% of visual test trials (see Figure 4). In contrast to the results from Experiment 1 (see Figure 2), only performance on the attended visual stream was significantly different from the performance of nonfamiliarized controls: attended visual, $t(18) = 3.67$, $p = .002$; unattended visual learning, $t(18) = 1.73$, $p = .10$; attended auditory, $t(18) = -1.81$, $p = .087$; unattended auditory, $t(18) = 0.85$, $p = .85$.

Effects of attention

Planned comparison of attended versus unattended performance within modality yielded a significant difference in the visual modality only:

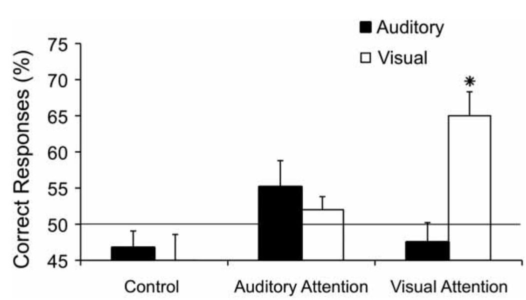


Figure 4. *Mean test performance (percentage correct out of 50) from Experiment 2. Visual and auditory ISL (implicit statistical learning) performance is presented for control, unattended, and attended conditions at slow presentation rate (750-ms stimulus onset asynchrony, SOA).*

attended versus unattended visual, $t(22) = 3.44$, $p = .002$; auditory, $t(22) = -1.65$, $p = .105$.

Modality effects at slower rates of presentation

As with the results in Experiment 1, data from the experimental groups were submitted to a repeated measures ANOVA. We find a main effect of modality, $F(1, 22) = 4.32, p = .050$, as well as a significant modality by attention interaction, $F(1, 22) = 8.98, p = .0066$. Along with evidence that learning is restricted to the attended visual stream, these results indicate a visual superiority in ISL at slow rates of presentation. That is, modality effects were obtained specifically when participants were devoting attention to each given input stream. While direct t tests of attended results across modalities of presentation reveal only a marginally significant result, $t(22) = 2.0$, $p = .058$, as with the results from Experiment 1, the interaction of modality and attention indicates that modality of presentation is not uniformly affecting learning across attentional conditions. Unlike Experiment 1, direct t tests indicate an opposite pattern of results of attention and learning: We find a significant effect of attention in the visual modality only and that significant learning is restricted to the attended visual stream. Together these results indicate that at the slower rate of presentation, visual ISL performed more robustly than auditory ISL. This is in contrast to the auditory superiority in ISL at fast presentation rates using the same materials and relative temporal dynamics as those in Experiment 1.

Comparing across rates of presentation

A direct comparison of results from Experiments 1 and 2 was conducted by submitting all experimental data to a three-way repeated measures ANOVA using between-subject factors of attended modality (auditory vs. visual) and rate (fast vs. slow) and within-subject factor of modality (auditory vs. visual). This analysis revealed a marginally significant main effect of modality, $F(1, 44) = 3.35$, $p = .074$, driven by slightly better performance across attended and unattended streams in the visual modality. In addition, we find a significant interaction of modality by attended modality, $F(1, 44) = 22.84$, $p < .002$. This interaction confirms that attention to a particular modality affects performance in the corresponding modality for both the visual and the auditory modalities. We also find a significant interaction between attended modality and rate of presentation, $F(1, 44) = 5.58$, $p = .023$. This result shows that the effect of rate is dependent on which modality is being attended. We do not obtain a rate by modality interaction, $F(1, 44) = 2.414$, $p = .127$, nor a modality by rate by attended modality interaction, $F(1, 44) = 0.010, p = .922$, which indicates that the rate manipulation does not affect one modality preferentially over the other (except when the modality is attended).

Taken together, the interactions of attended modality with rate and with modality of presentation provide support for our hypothesis that modality of presentation mediates differential performance across rates of presentation but this only occurs in the attended streams. This view is supported by the pattern of significant learning (i.e., greater performance than that of controls) in Experiments 1 and 2.

These results are broadly consistent with previous work that attention modulates ISL. However, they reveal a more complex relationship between attention and learning: Attending to a specific modality does in fact have an effect on ISL only in that modality—moreover, this modality-specific effect of attention interacts with rate of presentation. This pattern of results suggests not only that the impact of rate is not independent of attention but also that attention is not sufficient to overcome the modality-specific effects of rate.

Given the broad and complex effects of attention in the current task and our specific predictions that rate will interact with learning in the attended streams (see the introduction), we conducted separate analyses for attended and unattended performance to examine the effects of modality and rate within attentional condition. This analysis allows us to verify that the interactions discussed above are in fact driven by attended performance and not patterns of unattended learning. For each group (attended and unattended), we submitted the data to a repeated measures ANOVA (rate of presentation and within-subject factor: modality).

For the unattended conditions, we found a main effect of modality, $F(1, 44) = 4.80$, $p = .034$, but no effects of rate or interaction of rate and modality. Examination of the mean unattended performance across conditions reveals that visual unattended learning is superior to auditory unattended learning.

The same analysis of attended performance revealed a markedly different pattern of results. We find no main effect of modality or rate (Fs < 1) but a significant modality by rate interaction: $F(1, 44) = 6.47$, $p = .015$. These results confirm our predictions that modality and rate interact with learning but only when attention is deployed. These results also indicate that selective attention is not sufficient to overcome modality-specific differences in learning. Interestingly, at all rates, unattended visual performance is better than unattended auditory performance. By contrast, any effects of modality in the attended streams are mediated by rate of presentation.

Discussion

The perceptual literature predicts that the decrease in rate of presentation should have opposite effects on auditory and visual processing: A decrease in rate will disrupt auditory perceptual grouping while simultaneously easing the temporal processing in the visual modality. We find this differential pattern of performance in ISL, indicating that perceptual conditions are key to implicit learning ability. In the current experiment, rate of presentation is slowed to half the rate of Experiment 1. While auditory superiority is observed in the "fast-rate" experiment, we fail to observe any significant learning in the auditory modality at the slow rate of presentation. In stark contrast, we observe significant visual ISL at slower rates and a visual superiority effect.

As in Experiment 1 and previous research (Toro et al., 2005; Turk-Browne et al., 2005), we find that attention appears to aid ISL in general as exemplified by a modality by attention interaction. In addition, we report an attention by timing interaction across Experiments 1 and 2, suggesting that both attention and timing are independent factors affecting ISL ability. Previous research has suggested that attention is sufficient to overcome poor perceptual grouping conditions in the visual modality (Baker et al., 2004; Pacton & Perruchet, 2008). However, we fail to find evidence that attention is sufficient to overcome adverse perceptual grouping conditions in the current experiments. The next experiment further controls for the presence of unattended stimuli and any effects of cross-modal presentation in Experiment 2, in addition to providing a replication of the rate by modality interaction.

EXPERIMENT 3: INTERLEAVED BLANK STREAM

Finally, Experiment 3 addresses the possibility that the cross-modal interleaved design employed in Experiments 1 and 2 introduced cross-modal interference or decrements of selective attention due to the presentation of unattended stimuli from another modality (Spence & Driver, 1997). To this end, Experiments 3A (375-ms SOA) and 3B (750-ms SOA) use the same timings of the attended streams as those in Experiments 1 and 2, respectively (see Figure 3), but, instead of presenting unattended stimuli from the second input stream, pauses of equivalent length were presented. For example, if three unattended elements at 375-ms SOA were presented in Experiment 1, a pause of 1,125 ms (375 ms $\times$ 3) was presented in place of these unattended elements, as illustrated in Figure 1C. This manipulation preserved the temporal structure of the familiarization stream while removing any potential cross-modal interference and cost of dual-modality presentation. In addition, this experiment is meant to ameliorate any effects of the attentional manipulations employed in Experiments 1 and 2 and increases transitional probabilities of the familiarization stream. Without the unattended elements, transitional probabilities are equivalent to presenting the stream in isolation (see Table 1).

Method

The attended streams from Experiments 1 and 2 were used in the current experiment with the unattended elements removed and replaced with equal-length pauses in stimulus presentation. Thus the timing, materials, and methods were preserved from the previous two experiments with the exception of the removal of the unattended, cross-modal stream.

Participants
Another 32 participants (16 each for Experiments 3A and 3B) were recruited from introductory psychology classes at Cornell University to participate in exchange for course credit or \$10/hour.

Procedure
For each experiment, participants were randomly assigned to one of two conditions: nonfamiliarized controls or experimental groups. After being familiarized with the stimuli from one modality (with the other one "blanked out"), participants were then tested in that same modality with test trials for that modality only presented in random order. Then, participants went on to the familiarization, as described above, and testing trials for the other modality. Test trials were constructed using the same procedure as that in Experiments 1 and 2. Modality order was counterbalanced across participants. The same number of triplet groupings and interleaved sequence were employed. Because a single unimodal familiarization stream was presented during familiarization, there was no manipulation of selective attention.

Results and discussion

Experiment 3A
Participants in the nonfamiliarized control group responded correctly to 52% of visual and 51% of auditory test trials and did not perform significantly better than chance in either modality: visual, $t(7) = 0.97$, $p > .36$; auditory, $t(7) = 0.10$, $p > .92$. Participants in the experimental group responded correctly to 51% of the visual test trials and 64% of the auditory test trials. Only performance in the auditory modality was significantly better than that of nonfamiliarized controls: visual, $t(14) = 0.84$, $p = .41$; auditory, $t(14) = 2.33$, $p = .035$ (see Figure 5). We find a

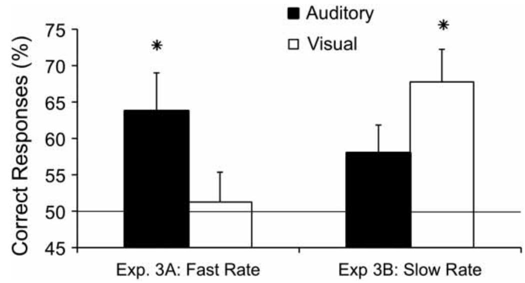


Figure 5. *Mean test performance for Experiment 3. Auditory and visual streams are presented with identical timing to that in Experiments 1 and 2 but without the unattended stimuli. Both modalities are attended and presented in counterbalanced order within participants. Left: Experiment 3A using the fast rate of presentation from Experiment 1. Right: Experiment 3B using the slow rate of presentation from Experiment 2.*

significant difference in mean performance across modalities, $t(15) = -4.79$, $p < .001$.

To verify that there were no effects of our within-subject design, we investigated possible order effects. Comparison of performance dependent on order of presentation revealed no effect of presentation order in either modality: visual, $t(6) = 0.131$, $p = .90$; auditory, $t(6) = -1.59$, $p = .16$.

Inspection of the left panel of Figure 5 readily reveals the replication of attended performance in Experiment 1. Statistical comparison between current experimental results and performance in attended conditions from Experiment 1 revealed no significant difference in either modality: visual, $t(14) = 0.83$, $p = .42$; auditory, $t(14) = -0.18$, $p = .86$. Thus, current results replicate attended performance in Experiment 1, demonstrating that the presence of unattended cross-modal stimuli has no effects on ISL performance in either the visual or the auditory modalities.

Together, the results from Experiments 1 and 3A demonstrate that auditory presentation yields greater learning performance than visual presentation. This finding is consistent with previous studies, all of which have employed similarly fast rates of presentation (Conway & Christiansen, 2005, 2009; Robinson & Sloutsky, 2007; Saffran, 2002).

Given that the interleaved experimental design entails that a single stream is temporally interrupted by the elements from the other stream, Experiments 1 and 3A also present evidence that auditory ISL is robust to short disruptions caused by unattended visual stimuli in Experiment 1 and pauses (equivalent length to visual stimuli in Experiment 1) in Experiment 3A. These pauses were on average 1.13 s in length, which is within the perceptual grouping tolerance of the auditory modality (Mates et al., 1994; see introduction).

Experiment 3B

Nonfamiliarized controls responded correctly to 51% of visual and auditory test trials and did not perform significantly better than chance in either modality: visual, $t(7) = 0.39$, $p = .71$; auditory, $t(7) = 0.56$, $p = .59$. Participants in the experimental group responded correctly to 68% of the visual test trials and 58% of the auditory test trials. In contrast to results from Experiment 3A, only performance in the visual modality differed significantly from that of nonfamiliarized controls: visual, $t(14) = 2.30$, $p = .037$; auditory, $t(14) = 1.66$, $p = .12$ (see Figure 5). We do not, however, find a significant difference in learning across modalities, $t < 1$.

As in Experiment 3A, we tested for possible order effects by comparing performance in each modality dependent on presentation order. We found no evidence of order effects for either modality: visual, $t(6) = 0.82$, $p = .45$; auditory, $t(6) = -0.62$, $p = .56$.

Since the purpose of the current experiment was to replicate the results from Experiment 2, performance in the attended streams from Experiment 2 was compared to that of the experimental group. This analysis revealed no significant difference in either modality: visual, $t(14) = -0.738$, $p = .47$; auditory, $t(14) = -0.54$, $p = .6$ (see Figure 5). Thus, along with Experiment 3A, we find no effect of the presence of unattended cross-modal stimuli on ISL across presentation rates.

Analyses comparing attended performance in Experiments 1 and 2 suggest that timing and modality interaction in ISL. To test for this pattern of results in the within-subject design of Experiments 3A and 3B, data from experimental groups were submitted to a repeated measures ANOVA (auditory and visual ISL; timing). This analysis revealed no main effect of modality, $F(1, 14) = 0.097$, $p = .76$. Crucially we find a significant modality by timing interaction, $F(1, 14) = 6.36$, $p = .024$. Together with the significant difference in auditory and visual performance in Experiment 3A, these results confirm the earlier result that presentation rate is a significant factor modulating ISL and reveals that timing has a differential effect on ISL across visual and auditory modalities in two different experimental paradigms: with and without attentional manipulations to a particular modality of presentation. While the significant interactions of modality and timing or rate of presentation in Experiments 3A and 3B replicate the effects found in Experiments 1 and 2, in both of these experiments the interaction of rate and modality is found across participants and without random sampling across these experimental groups (i.e., there is only random sampling within Experiment 1).

GENERAL DISCUSSION

In three experiments, participants were presented with auditory and visual statistical regularities under different timing conditions. While statistical information remained the same, performance was not equivalent across perceptual modalities in these different rates of presentation. In attended modalities, perceptual modality and rate of presentation interact to affect learning: At fast rates of presentation, similar to previous studies, auditory ISL performed better than visual; however, at a slower presentation rate, auditory ISL was reduced, and visual ISL became superior. This effect is replicated in Experiment 3, in which the timing conditions are kept constant but streams are presented without unattended stimuli. Thus, the current experiments have uncovered a new phenomenon in implicit statistical learning where timing changes result in differential effects on ISL across visual and auditory modalities.

These learning results are consistent with well-known perceptual differences across visual and auditory modalities. As reviewed in the introduction, visual and auditory modalities appear to process time and space differently. Specifically, perceptual performance across modalities can be characterized by the visual:spatial::auditory:temporal analogy, where the visual modality preferentially processes spatially arrayed information while the auditory modality processes temporal information more robustly than the visual modality. While it is clear that under some circumstances this characterization does not apply (e.g., Potter, 1976), in the current type of perceptual task, these modality-differences are well established. While this perceptual effect does not limit the ability to recognize single objects, spatial or temporal presentation of multiple stimuli does affect processing across visual and auditory sensory modalities.

Our learning results parallel this well-known effect in the perceptual literature: Visual learning is increased with a slower rate while auditory processing is decreased. This behavioural phenomenon along with numerous others (e.g., Conway & Christiansen, 2005, 2006; Robinson & Sloutsky, 2007; Saffran, 2001) has established that ISL is not equivalent in the visual and auditory sensory modalities, even though the same statistical information is presented, and all stimuli are clearly presented in all perceptual conditions.

We also investigated the effect of attention to a sensory modality across both timing conditions. While it is generally accepted that attention is a significant modulatory factor aiding in both visual and auditory ISL (visual: Turk-Browne et al., 2005; auditory: Toro et al., 2005), it is unclear whether attention is necessary for learning to take place (e.g., Saffran et al., 1997). We do not find any evidence for unattended learning (in Experiments 1 and 2). Therefore, we corroborate previous research showing that attention significantly boosts ISL in both visual and auditory sensory modalities and may be necessary for learning.

Turning to results for attended modalities, previous research on perceptual grouping effects in ISL has emphasized the relationship between perceptual conditions and attention. Baker et al. (2004) point to the "automatic spreading of attention induced by perceptual-grouping" (p. 465) as the mechanism by which perceptual grouping aids ISL. Under unfavourable grouping conditions, it has been argued that ISL can occur if the relevant stimuli are attended (Baker et al., 2004; Pacton & Perruchet, 2008). This approach emphasizes perceptual grouping as a factor that modulates attention, which in turn facilitates ISL, as opposed to treating perceptual grouping as a separate modality-specific factor affecting learning from environmental regularities.

If it were the case that attention is sufficient to compensate for poor perceptual conditions, we should observe equivalent learning in all attended streams regardless of modality and rate of presentation. The current results do not support this view: We observed a modality-specific decrement in ISL under disruptive perceptual grouping conditions even when there are no competing demands for attention. Thus, we find evidence that attention cannot always compensate for poor perceptual conditions.

Our results establish that favourable perceptual conditions and selective attention to a particular modality may both be required in order for participants to learn from environmental regularities. In the current learning paradigm, attention appears to be a prerequisite for learning: There is no learning in any unattended stream regardless of perceptual modality or timing condition. However, selective attention to a particular modality is not sufficient for learning. In fact, even with attention, perceptual grouping conditions and modality of presentation interact to determine whether or not participants learn. Thus, we find that attention and perceptual conditions amenable to a particular modality are both necessary for ISL.

Overall, these results are consistent with ISL being mediated by mechanisms that are sensitive to the *perceptual nature* of the input in addition to its statistical structure (e.g., transitional probabilities and frequencies). Recent neuroimaging evidence has supported just such a scenario: Turk-Browne, Scholl, Chun, and Johnson (2009) report increased visual cortex activity during the observation of learnable visual sequences demonstrating that sensory cortices are probably involved in learning the underlying structure from visual statistics. Thus, both behavioural and neuroimaging results have suggested that ISL is sensitive to perceptual processes.

Given these modality-specific learning effects, any mechanisms for ISL must be able to account for differences in learning across sensory modalities. Here we consider three types of mechanisms for their ability to accommodate modality-specific differences in learning as well as the domain- and modality-general quality of ISL. In Figure 6, we present a simplified characterization of these possible architectures. Throughout the paper, we have been referring to ISL as a behavioural phenomenon in which exposure to statistical regularities affects subsequent behaviour. By contrast, these candidate characterizations present types of possible mechanisms that could underlie this behaviour: Specifically, we consider different ways in which perceptual systems—responsible for the recognition and understanding of sensory information—and learning mechanisms—the acquisition of knowledge as a result of experience—could contribute to ISL ability.

As asserted in the introduction, it is difficult for the *standard view* of the mechanisms underlying ISL to account for these modality-specific patterns of learning. According to this view, ISL involves a single undifferentiated mechanism for which the nature of the input beyond its statistical characteristics is irrelevant (e.g., Perruchet & Pacton, 2006; Reber, 1989; Shanks et al., 1997). In other words, visual and auditory perception are separate unimodal processes that feed into a single learning mechanism. Consistent with this type of architecture, many prominent models make an "equivalence assumption" that the same statistical information presented across modalities should result in equivalent learning. Thus, while knowledge gained from ISL may be specific to the input stimuli (i.e., learning has limited or no transfer to other stimuli), the learning mechanism is not affected by the perceptual nature of the input (e.g., auditory stimuli are learned equivalently to visual stimuli). This standard architecture is consistent with the general modality- and domain-general nature of ISL given that there is a single learning mechanism that receives input from both

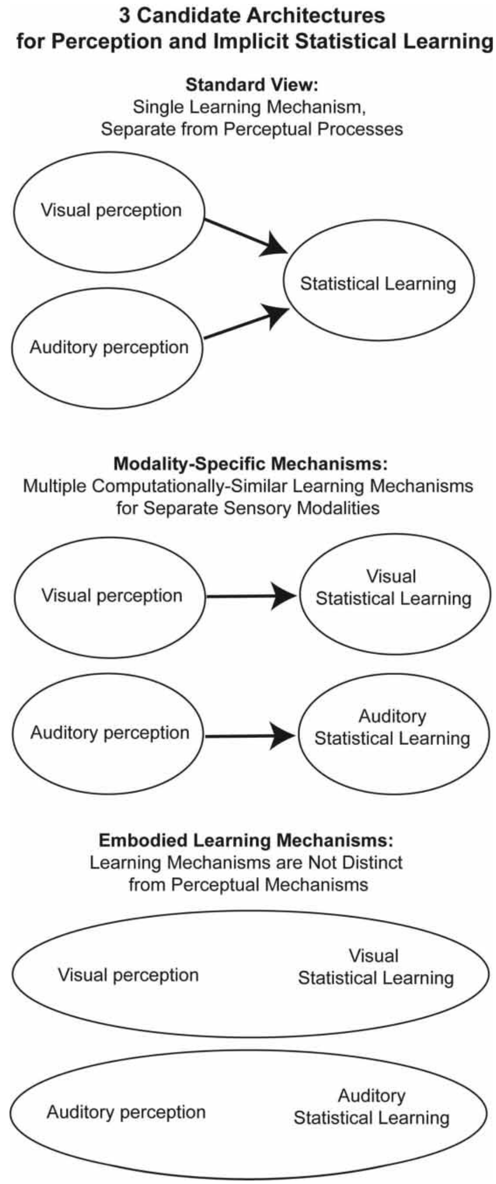


Figure 6. *Simplified characterization of three possible architectures for perception and ISL (implicit statistical learning): The top architecture is the* standard view *in the literature where perception (visual and auditory) is a separate process that feeds into a single, general learning mechanism. The middle architecture is a* modality-specific *architecture with separate but computationally similar learning mechanisms for both visual and auditory perception, but perception and learning are still distinct processes. At the bottom, we present an* embodied *architecture where perception and learning are not distinct processes, but learning mechanisms are grounded in perceptual processing.*

perceptual systems. However, in our view, it is unclear how a learning mechanism that focuses on statistical information to the exclusion of the perceptual nature of the input would be able to accommodate the differential learning effects we report across modalities. One possibility is that the unimodal perceptual systems are tasked with detecting perceptual "units", while the singular learning mechanism can determine and track the statistical regularities of these units, and differences in timing or other perceptual conditions make resolution of "units" in a stream of stimuli more difficult. The most obvious definition of a perceptual unit is a single object. However, in the current study and previous work, the perception of individual perceptual objects is the same across modalities.

We present two alternative types of architecture that, we believe, more easily accommodate the modality-specific patterns of learning observed in the current ISL paradigm. The first is a *modality-specific* architecture consisting of separate but computationally similar learning mechanisms for both visual and auditory perception, thus allowing perception and learning to remain distinct processes. The multiple, modality-specific learning mechanisms, which characterize this type of architecture, make it possible for differences in learning across sensory modalities to emerge in the learning mechanisms themselves (e.g., by detecting types of patterns similar to previous experience, spatial patterns in the visual system), as well as accounting for the modality- and domain-general nature of ISL. Two disadvantages to this architecture are that it is unparsimonious as it includes an additional, if computationally similar, learning mechanism and does not explain cross-modal learning effects (e.g., Robinson & Sloutsky, 2007).

In an *embodied* architecture, perception and learning are not distinct processes but the learning mechanism is a part of, or embodied in, perceptual processing. Recent work has suggested that perception is a prediction-based process (e.g., Summerfield, Trittshuh, Monti, Mesulam, & Enger, 2008) where the likelihood of a given stimulus affects perceptual processing. In this view of perception, sensitivity to statistical information is already a property of visual and auditory perception. Thus, changes in perception as a result of statistical information (e.g., identification of multiple objects as a single triplet) might only be a modification from the usual perceptual process. Like both the modality-specific and standard architectures, the embodied view is able to accommodate the modality- and domain-general nature of ISL. An embodied architecture provides a clear way to accommodate the modality-specific learning effects based upon perceptual conditions such as perceptual grouping: Because learning is embodied in perceptual processing, perceptual biases that are seen in perceptual tasks will robustly transfer to the process of acquiring knowledge through exposure to statistical regularities. Unlike a modality-specific architecture, an embodied mechanism is more parsimonious; it may also better accommodate cross-modal effects in ISL (e.g., learning based on cross-modal statistical regularities could be embodied in multisensory perceptual systems).

In sum, the aim of the current paper is not to conclusively support one type of ISL mechanism but to further elucidate the importance of perceptual processing in learning from statistical regularities. We have highlighted the ways in which these three types of architecture could accommodate modality-specific patterns of learning as reported in the current paper while maintaining modality- and domain-general learning. However, we assert that both the modality-specific and the embodied view more readily support differential learning across the visual and auditory modalities.

A further important question is to understand the origins of these modality-specific learning differences. For instance, are the perceptual constraints observed here true for all categories of sound stimuli or just for speech sounds? Similarly, is it possible that auditory ISL is more temporally tuned because of language-specific experience? Alternatively, is hearing temporally biased due to generic sensory and/or perceptual factors apart from experience with spoken language?

Our understanding of ISL as an important mechanism in cognition and development has progressed from it being characterized as language-specific, to domain-general and abstract, to current thinking emphasizing the effects of perceptual, attentional, and modality-specific constraints. The evidence appears to disconfirm the idealized conception of ISL as a single, undifferentiated mechanism that operates apart from other perceptual and cognitive constraints. By recognizing and further discovering the complexities governing and affecting the operation of this ubiquitous learning mechanism, we may better understand fundamental processes of language, development, and cognition.

Original manuscript received 13 November 2009
Accepted revision received 28 October 2010
First published online 22 February 2011

REFERENCES

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21,* 899–912.

Audacity Team. (2005). Audacity (Version 1.2.2) [computer software]. Retrieved December 1, 2005, from http://audacity.sourceforge.net/

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15,* 460–466.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7,* 84–91.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, *21,* 515–532.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, & 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.

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21,* 561–580.

Fiser, J., & Aslin, R. A. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12,* 499–504.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20,* 1–55.

Handel, S., Weaver, M. S., & Lawson, G. (1983). Effect of rhythmic grouping on stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *9,* 637–651.

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

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance*, *14,* 318–320.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35,* 71–98.

Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, *23,* 1369–1386.

Mates, J., Radil, T., Müller, U., & Pöppel, E. (1994). Temporal integration in sensorimotor synchronization. *Journal of Cognitive Neuroscience*, *6,* 332–340.

Pacton, S., & Perruchet, P. (2008). An attentional-based account of adjacent and non-adjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *1,* 80–96.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1,* 29–55.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17,* 398–422.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: Two approaches, one phenomenon. *Trends in Cognitive Sciences*, *10,* 233–238.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2,* 509–522.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118,* 219–235.

Robinson, C. W., & Sloutsky, V. M. (2007). Visual statistical learning: Getting some help from the auditory modality. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 611–616). Austin, TX: Cognitive Science Society.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44,* 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47,* 172–196.

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 human infants and adults. *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.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, *50A,* 216–252.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106,* 1558–1568.

Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59,* 389–403.

Summerfield, C., Trittshuh, E. H., Monti, J. M., Mesulam, M.-M., & Enger, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11,* 1004–1006.

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

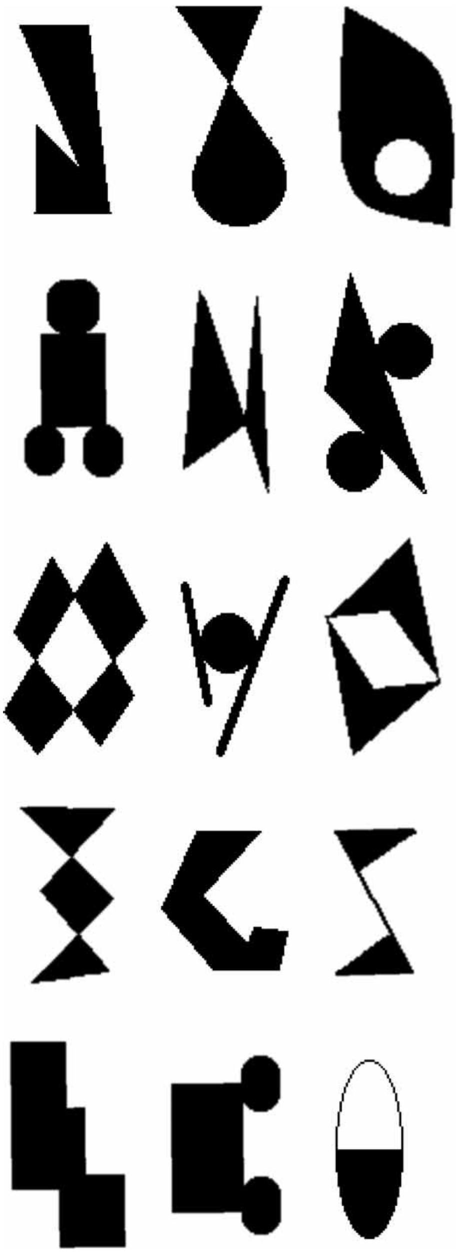
Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134,* 552–564.

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,* 1934–1945.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). New York, NY: Harcourt.

APPENDIX A

Fifteen shapes used in all experiments, grouped into arbitrary triplets



APPENDIX B

Monosyllabic nonwords used as auditory stimuli in all experiments

The 225-ms monosyllabic nonwords used in Experiments 1 and 3A

bu, cha, da, el, feng, jic, leep, rau, roo, rud, sa, ser, ta, wif, zet

The 450-ms monosyllabic nonwords used in Experiments 2 and 3B

bu, cha, dak, eeg, feng, jeen, jic, meep, pel, rauk, rous, rud, sa, ser, wif