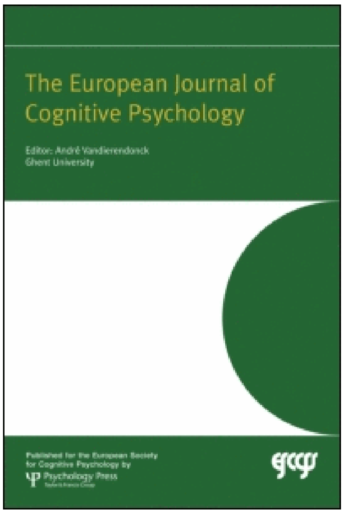

European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning

Christopher M. Conway [a]; Morten H. Christiansen [b]
[a] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA [b] Department of Psychology, Cornell University, Ithaca, NY, USA

First Published on: 01 July 2008

To cite this Article Conway, Christopher M. and Christiansen, Morten H.(2008)'Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning',European Journal of Cognitive Psychology,21:4,561 — 580

To link to this Article: DOI: 10.1080/09541440802097951

URL: http://dx.doi.org/10.1080/09541440802097951

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2009, 21 (4), 561–580

Psychology Press
Taylor & Francis Group

Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning

Christopher M. Conway

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Morten H. Christiansen

Department of Psychology, Cornell University, Ithaca, NY, USA

Across a wide range of tasks, vision appears to process input best when it is spatially rather than temporally distributed, whereas audition is the opposite. Here we explored whether such modality constraints also affect implicit statistical learning in an artificial grammar learning task. Participants were exposed to statistically governed input sequences and then tested on their ability to classify novel items. We explored three types of presentation formats—visual input distributed spatially, visual input distributed temporally, auditory input distributed temporally—and two rates of presentation: moderate (4 elements/second) and fast (8 elements/second). Overall, learning abilities were best for visual-spatial and auditory input. Additionally, at the faster presentation rate, performance declined only for the visual-temporal condition. Finally, auditory learning was mediated by increased sensitivity to the endings of input sequences, whereas vision was most sensitive to the beginnings of sequences. These results suggest that statistical learning for sequential and spatial patterns proceeds differently across the visual and auditory modalities.

Keywords: Implicit learning; Statistical learning; Vision; Audition; Modality effects.

Correspondence should be addressed to Christopher M. Conway, Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th St., Bloomington, IN 47405, USA. E-mail: cmconway@indiana.edu

This work was done as part of the first author's doctoral dissertation while at Cornell University. Partial preparation of this manuscript was supported while on a postdoctoral training grant from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant DC-00012; program director, David B. Pisoni). Earlier versions of this research were presented at the 43rd and the 47th annual meetings of the Psychonomic Society.

http://www.psypress.com/ecp DOI: 10.1080/09541440802097951

Intuitively, each sensory modality seems biased to handle particular aspects of environmental input. For instance, vision and audition appear to be most adept at processing spatial and temporal input, respectively (Kubovy, 1988). Consider that it takes very little time—often a single glance—in order to recognise a relatively complex visual scene, but it generally takes several seconds or longer to recognise a melody or voice. That is, a brief "snapshot" of sound is fairly incomprehensible, whereas a visual snapshot may be quite coherent. This somewhat anecdotal evidence, bolstered by studies of perception and memory, suggests that in visual cognition, the dimensions of space weigh most heavily, whereas for audition, the temporal dimension is most prominent (Freides, 1974; Geldard, 1970; Kubovy, 1988; O'Connor & Hermelin, 1978; Penney, 1989).

These modality constraints have been proposed to affect the manner in which stimuli are perceived (Mahar, Mackenzie, & McNicol, 1994; Repp & Penel, 2002), maintained in short-term memory (Collier & Logan, 2000; Gardiner & Cowan, 2003; Glenberg & Swanson, 1986; Penney, 1989), and learned (Conway & Christiansen, 2005; Handel & Buffardi, 1969; Saffran, 2002). Even so, researchers do not always give modality effects due attention. In fact, it has been suggested that cognitive psychologists ought not to consider the senses as being separate entities at all (e.g., Ghazanfar & Schroeder, 2006; Marks, 1978; Stoffregen & Bardy, 2001). Or at the very least, modality effects ought to be deemphasised, with a focus instead on the importance of amodal (e.g., Gibson, 1966) information in the environment that can be detected by any of the sense modalities. Unfortunately, these views tend to obfuscate the possible presence of modality differences that may affect processing, which in turn can hinder the development of a complete theory of human cognition.

Effects of presentation modality are still largely unexplored in the realm of implicit statistical learning, defined as the ability to extract environmental regularities through automatic learning mechanisms operating outside of immediate awareness (cf. Cleeremans, Destrebecqz, & Boyer, 1998; Conway & Christiansen, 2006; Perruchet & Pacton, 2006; Reber, 1967; Saffran, Aslin, & Newport, 1996; Stadler & Frensch, 1998). Statistical learning is a fundamental ability believed to underlie important aspects of language, cognition, and perception (Altmann, 2002), including speech segmentation (Saffran, Newport, & Aslin 1996), spoken language processing under degraded listening conditions (Conway, Karpicke, & Pisoni, 2007), the learning of orthographic regularities in printed language (Pacton, Perruchet, Fayol, & Cleeremans, 2001), visual scene perception (Fiser & Aslin, 2001; Olson & Chun, 2002), tactile pattern processing (Conway & Christiansen, 2005), and visual-motor skill acquisition (Cleeremans & McClelland, 1991). Because implicit statistical learning appears to make contact with so many different cognitive domains, understanding its underlying mechanisms and constraints is an important research goal.

However, despite the interest in statistical learning, very few studies have directly compared learning across sensory domains; fewer still have investigated to what extent spatial and temporal constraints affect statistical learning. One exception is Conway and Christiansen (2005), who examined statistical learning of tactile, visual, and auditory input sequences and found that auditory learning of statistical regularities exceeded both visual and tactile learning. Their work followed previous suggestions that visual implicit learning proceeds better when the stimulus elements are presented simultaneously—that is, spatially arrayed—rather than sequentially, presumably because a simultaneous format permits better chunking of the stimulus elements (Gomez, 1997; Saffran, 2002). In addition, using the contextual-cueing paradigm, which is distinct but related to statistical learning, it has been found that spatial features play a more important cueing role than surface features such as colour in the detection of invariant visual patterns (e.g., Olson & Chun, 2002). Finally, relative spatial arrangements appear to be automatically extracted by participants in a visual statistical learning task (Conway, Goldstone, & Christiansen, 2007). Together, these studies suggest an intimate connection between visual implicit learning and spatially arrayed patterns. However, to date there has not been a carefully controlled study investigating how temporal and spatial constraints interact with visual and auditory statistical learning.

In this report, we directly test the hypothesis that temporal and spatial constraints differentially affect visual and auditory learning of statistical patterns. Using the AGL paradigm, we expose participants to statistically governed visual or auditory input sequences generated from an artificial grammar and then test learners on their ability to generalise their knowledge to novel patterns. Our primary manipulation is the manner in which the visual input is distributed: spatially or temporally. To serve as a comparison group, we also include an auditory condition with input distributed temporally. In line with a modality-constrained view, we predict that learning will be greatest for the visual-spatial input, comparable to the auditory (temporal) condition, and poorest for the visual-temporal format. An additional aim of this study is to explore the effect that rate of presentation has on learning, an issue that has not been explored in full for statistical learning tasks (but see Frensch & Miner, 1994; Turk-Browne, Jungé, & Scholl, 2005). Faster presentation rates may magnify the effect of modality constraints by adversely affecting learning in the nonpreferred mode of processing (see Collier & Logan, 2000, for an example in the realm of short-term memory). Consequently, we predict that at the fast rate, learning in the visual-temporal condition will show the largest decrement in performance.

Before detailing the experiment, we first briefly review relevant AGL work that has investigated the general issue of how input format affects learning. One line of work has purportedly shown that participants can transfer their implicit knowledge of the grammatical regularities from one letter vocabulary (e.g., M, R, T, V, X) to another (e.g., N, P, S, W, Z) (Brooks & Vokey, 1991; Mathews et al., 1989; Reber, 1969; Shanks, Johnstone, & Staggs, 1997) and even across sense modalities (Altmann, Dienes, & Goode, 1995). This suggests that what is learned from statistical learning is independent of the perceptual features of the stimuli, possibly some form of abstract or amodal knowledge. However, other studies have called the transfer results into question (Redington & Chater, 1996). For instance, Tunney and Altmann (1999) showed that when stimuli are controlled so that ungrammatical test items do not differ in terms of repetition patterns or starting element legality, participants are not able to successfully apply grammatical knowledge of training stimuli to a different sensory modality. This suggests that rather than abstract knowledge, at least part of what is learned is modality or stimulus specific.

A second line of work has investigated how differences in stimulus format affect learning. In some cases, changing particular aspects of the stimuli between the exposure and test phases appears to affect AGL performance. For instance, Chang and Knowlton (2004) showed that changing the font affected how well participants could learn the statistical patterns in the letter strings, suggesting that their knowledge was at least partly based on stimulus-specific features of the input. On the other hand, Pothos and Bailey (2000) compared AGL for embedded geometric shapes versus geometric shapes presented in sequence, and found no differences (see also, Pothos, Chater, & Ziori, 2006). In sum, the existing data is unclear regarding the specific role that input type and sensory modality play in constraining learning; specifically, how temporal and spatial constraints interact with visual and auditory implicit statistical learning has been relatively unexplored.

METHOD

Subjects

One hundred and forty-four subjects (twelve in each of twelve conditions) with normal hearing and normal or corrected-to-normal vision were recruited from psychology classes at Cornell University, earning course credit for their participation. Data from two subjects were discarded, one due to the participant being colour blind and the other due to experimenter error.

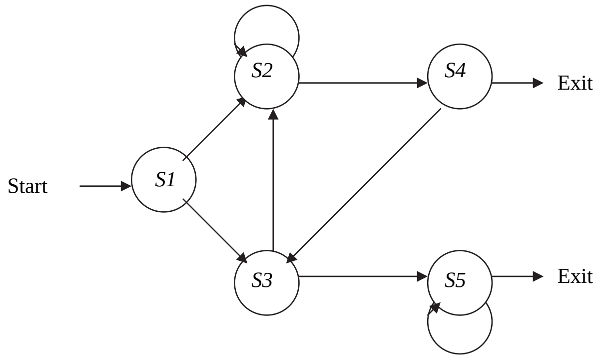

Figure 1. Artificial grammar used in the experiment. Each numeral was mapped onto a particular auditory or visual stimulus element, depending on the condition.

Materials

Figure 1 shows the artificial grammar, consisting of five unique elements, each represented by a numeral (1–5). Each numeral was mapped onto elements from one of four vocabularies, corresponding to three input conditions (auditory, visual-temporal, and visual-spatial).

For the acquisition phase we used 12 legal sequences that were generated from the grammar.[1] Each legal sequence was used twice to create a set of 12 learning pairs (see Appendix A). Six of the pairs consisted of the same sequence presented twice (matched pairs), whereas the other six pairs consisted of two different sequences (mismatched pairs). These matched and mismatched pairs were used in conjunction with a same–different judgement task.

The test set (see Appendix B) consisted of 10 novel legal and 10 illegal sequences. Legal sequences were produced from the finite-state grammar. The illegal sequences each begin with a legal element (i.e., 1 or 4), followed by one or more illegal transitions and ending with a legal element (i.e., 2, 3, or 5). For example, the illegal sequence 1-4-5-1-3-3 begins and ends with legal elements (1 and 3, respectively) but contains several illegal interior transitions not allowed by the grammar (1-4, 4-5, 5-1, and 3-3). The legal and illegal sequences can be described as differing from one another in terms of cooccurrence statistics of adjacent elements. That is, a statistical learning mechanism could discern which novel test sequences are illegal by noting the presence of pairwise element combinations that did not occur in the training set.

[1] See Meulemans and van der Linden (1997) for discussion regarding how the number of training/acquisition items may affect the nature of learning.

Each of the five stimulus elements (1–5) were mapped onto one of three vocabularies, corresponding to three types of modality/format. Importantly, the timing of all stimuli was equated across these four conditions.

Visual-temporal. For the visual-temporal conditions, the stimulus elements consisted of different coloured squares (1 = red, 2 = blue, 3 = yellow, 4 = green, 5 = black) appearing sequentially in the centre of the computer screen at approximately eye level. Each square (2.6×2.6 cm) appeared for 250 ms in the "slow" input condition and 125 ms in the "fast" condition (with no pauses between the presentation of each square in a sequence). Thus, the sequence 1-2-1-3 consists of a temporal sequence of red, blue, red, and yellow squares (see Figure 2).

Visual-spatial. For the visual-spatial conditions, the stimulus elements consisted of the same coloured squares described earlier, except that all squares in a sequence were presented simultaneously along a horizontal row, from left to right, at approximately eye level. The timing of the stimuli was equal to the cumulative presentation time of the visual-temporal stimuli. That is, in the slow condition, a visual-spatial sequence was displayed for a number of milliseconds equal to $250 \times N$, where N is the number of squares in the sequence. For the fast condition, a sequence was displayed for $125 \times N$ ms. Thus, the visual-spatial sequence 1-2-1-3 consists of a row of squares appearing simultaneously for $(250 \times 4) = 1000$ ms or $(125 \times 4) = 500$ ms, from left to right: red, blue, red, yellow (see Figure 3).

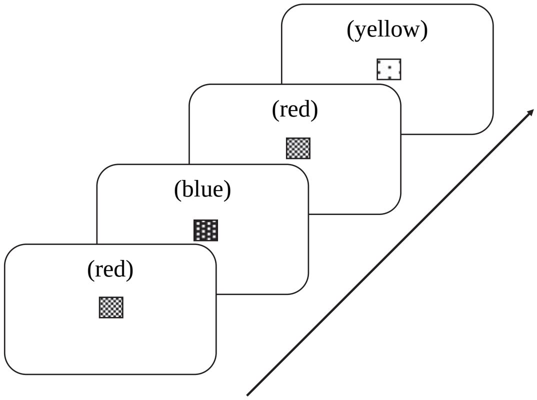

Figure 2. Example of visual-temporal sequence 1-2-1-3. The arrow designates the flow of time. Note that participants did not see the labels for each colour; they are included here only for illustrative purposes.

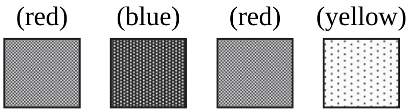

Figure 3. Example of visual-spatial sequence 1-2-1-3. Note that participants did not see the labels for each colour; they are included here only for illustrative purposes.

Auditory. For the auditory conditions, the stimulus elements consisted of pure tones of various frequencies (1 = 261.6 Hz, 2 = 277.2 Hz, 3 = 349.2 Hz, 4 = 370 Hz, and 5 = 493.9 Hz) corresponding to musical notes C, C#, F, F#, and B, respectively, played through headphones. For the "slow" and "fast" input conditions, each stimulus element (tone) had durations of 250 ms and 125 ms, respectively. As an example, the sequence 1-2-1-3 consists of the following four notes in this order: C, C#, C, and F.

Procedure

Participants were assigned randomly to one of 12 conditions, six control groups and six experimental groups. Within the experimental groups, the 3×2 design consisted of two factors: modality/format (visual-temporal, visual-spatial, and auditory,) and presentation rate (slow and fast). The slow and fast groups received the input material at a presentation rate of four and eight elements per second, respectively (corresponding to individual element durations of 250 ms and 125 ms). The six control groups were used as a baseline measure of performance. Consequently, the experimental groups received both acquisition and test phases whereas the control groups received the test phase only.

At the beginning of the acquisition phase, the experimental group participants were instructed that they would hear or see pairs of sequences. For each pair of sequences (listed in Appendix A), they had to decide whether the two sequences were the same or not and indicate their decision by pressing a button marked *YES* or *NO*. This match–mismatch paradigm, also used by Conway and Christiansen (2005), served as a way to encourage participants to pay attention to the stimuli without giving them explicit instruction that the sequences conformed to an underlying structure. Note that there is no feedback given during the task.

Each pair was presented six times in random order for a total of 72 exposures, with the timing parameters described earlier. In all conditions, a 2 s pause occurred between the two sequences of each pair and following the last sequence of the pair. After exposure to the two sequences, a prompt was displayed on the computer monitor asking for the participant's response, until a button press was made. After another 2 s pause, the next training pair was presented. The entire training phase in each condition lasted roughly 10 min for each participant.

Before the test phase, participants in the experimental group were told that the sequences they had just observed had been generated by a computer program that determined the order of the stimuli by using a complex set of rules. They were told that they would now be presented with new sequences. Some of these would be generated by the same program, whereas others would not. Participants were instructed to classify each new sequence according to whether or not they thought it was generated by the same rules as before, using two buttons marked *YES* and *NO*, and without feedback. The control participants, who did not participate in the acquisition phase, received an identical test task.

The 20 test sequences were presented one at a time, in random order, to each participant. The timing of the test sequences was the same as that used during the acquisition phase (250 ms or 125 ms element durations and 2 s pauses before and after each sequence).

RESULTS AND DISCUSSION

Acquisition-phase performance for the six experimental groups is shown in Table 1, which displays the mean number of correct match/mismatch decisions out of 72. We submitted the data to a two-way ANOVA with the factors modality/format and presentation rate. There was a main effect of modality/format, $F(2, 72) = 23.5$, $p < .0001$, but no main effect of presentation rate, $F(1, 72) = 0.11$, $p = .74$, nor a significant interaction, $F(2, 72) = 1.1$, $p = .34$. We conducted post hoc tests, collapsing across presentation rate, in order to compare performance among the three modality/format groups. To minimise Type I errors, Bonferroni corrections were applied to give a new alpha level of ($.05/3 = .016$). The results showed that acquisition phase performance for both the auditory and visual-spatial groups were

TABLE 1
Experimental group acquisition phase (match–mismatch) results

	Slow (4 elements/s)			*Fast (8 elements/s)*			*Comparison*
Modality/format	*M*	*SE*	*t(11)*	*M*	*SE*	*t(11)*	*t(22)*
Visual-temporal	61.3	1.92	13.2*	59.4	1.78	13.1*	0.73
Visual-spatial	65.8	1.06	28.2*	67.6	1.14	27.6*	−1.12
Auditory	68.7	1.01	32.4*	70.0	0.79	43.1*	−0.97

Mean values reported out of 72 possible correct. *T*-tests are conducted with respect to chance levels (36 out of 72).

*$p < .001$.

TABLE 2
Control group (baseline) test results

	Slow (4 elements/s)			*Fast (8 elements/s)*		
Modality/format	*M*	*SE*	*t(11)*	*M*	*SE*	*t(11)*
Visual-temporal	9.9	0.53	−0.16	8.9	0.65	−1.68
Visual-spatial	9.3	0.69	−0.97	9.25	0.54	−1.39
Auditory	10.8	0.47	1.76	10.75	0.48	1.57

Mean values reported out of 20 possible correct. *T*-tests are conducted with respect to chance levels (10 out of 20).

significantly greater than performance for the visual-temporal group, ($p <$ .001 for both). Furthermore, all six groups performed the match/mismatch task at greater than chance levels (see Table 1). Finally, we also contrasted performance for fast versus slow rates, for each of the three modality/format conditions separately. None of the contrasts were statistically significant (see Table 1), confirming that rate did not affect performance on the match/mismatch task for any of the modalities.

The control group test results are shown in Table 2, which displays the mean number of correct classification responses out of 20. As expected, none of the six groups performed different from chance levels, thus indicating that any learning displayed by the experimental groups is due to exposure to the material during the training phase.

The experimental group test results are displayed in Table 3. We submitted these data to a two-way ANOVA with the factors modality/format and presentation rate. There was a main effect of modality/format, $F(2, 72) = 7.65$, $p < .001$, but no main effect of presentation rate, $F(1, 72) = 0.97$, $p = .33$, nor a significant interaction, $F(2, 72) = 1.13$, $p = .33$.

TABLE 3
Experimental group test results

	Slow (4 elements/s)			*Fast (8 elements/s)*			*Comparison*
Modality/format	*M*	*SE*	*t(11)*	*M*	*SE*	*t(11)*	*t(22)*
Visual-temporal	12.2	0.68	3.17**	10.4	0.45	0.92	2.14*
Visual-spatial	13.1	0.91	3.39**	12.7	0.82	3.25**	0.34
Auditory	13.9	0.72	5.42***	14.3	0.69	6.29***	−0.42

Mean values reported out of 20 possible correct. *T*-tests for each of the six groups are conducted with respect to chance levels (10 out of 20). The comparison *t*-test compares performance for the slow and fast conditions for each modality/format.

*p <.05, **p <.01, ***p <.001.

We continued with several planned post-hoc comparisons. First, we compared performance among the three different modality/formats, collapsed across presentation rate, using a corrected alpha level of $(.05/3 = .016)$. The auditory group performance was greater than visual-temporal performance ($p < .001$). It is also relevant to note that visual-spatial performance was also greater than visual-temporal performance, using the noncorrected level of .05 ($p = .033$).

We also contrasted performance for fast versus slow rates, for each of the three modality/format conditions. The only significant result was for the visual-temporal group, $t(22) = 3.75$, $p < .05$, indicating that visual-temporal performance was significantly worse at the fast rate compared to the slow rate. Consistent with this finding, only the visual-temporal (fast) group performed the test task at chance levels (see Table 3).

In sum, these analyses reveal two key findings that can be easily appreciated by viewing Figure 4. First, all three groups demonstrated learning at the slow presentation rate, at levels comparable to one another. Second, only the auditory and the visual-spatial groups showed learning at the fast rate; the visual-temporal group performed much worse, in fact no

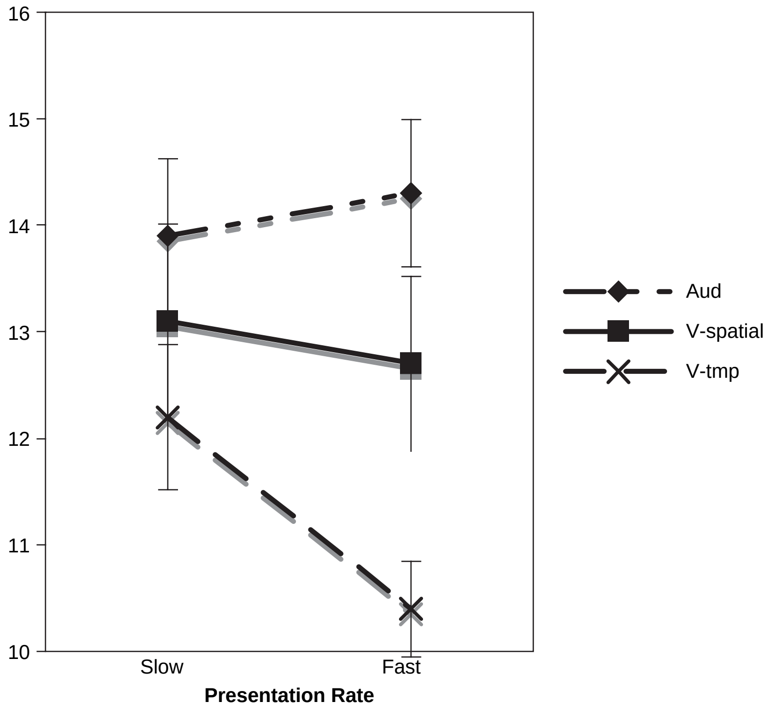

Figure 4. Experimental group mean test scores (for slow and fast rates) with error bars included.

better than chance levels. These findings suggest that the rate of presentation differentially affected learning in the three modality/format conditions. Thus, the ability to implicitly acquire visual-temporal statistical patterns is hindered severely at a fast rate of presentation, whereas statistical learning for auditory and visual-spatial patterns is not negatively impacted at fast presentation rates, at least for the rates used here.

Contribution of acquisition phase performance to test performance

One possible explanation of the quantitative modality differences is that there could be an association between acquisition-phase performance and test-phase performance. That is, perhaps better test-phase performance for particular modality/rate conditions merely was due to certain stimuli being more easily perceived and/or remembered (as assessed in the acquisition phase match/mismatch task). This may be possible, given that we found a significant effect of modality/format on the match/mismatch task. That is, the stimuli used in certain experimental conditions may have been more easily perceived in general which could directly affect the ease in which the statistical regularities are learned.

On the other hand, for the match/mismatch task, there was no decrement in performance at the fast presentation rate for any of the three input conditions, whereas there was a clear drop in test performance at the fast presentation rate for the visual-temporal condition. For this reason it is less likely that the test-phase results can be solely attributed to acquisition-phase performance.

In order to help tease apart these issues, we investigated the contribution of acquisition-phase performance to test-phase performance by computing the Pearson correlation coefficients between the acquisition and test phase for each of the three modality/format conditions separately (collapsed across presentation rate). The results revealed that only the visual-spatial correlation coefficient was statistically significant, $r = .50$, $p < .05$, whereas the other two coefficients were smaller and nonsignificant (auditory: $r = .28$, $p = .19$, visual-temporal: $r = -.002$, $p = .99$). These correlation analyses suggest that there may in fact be an association between acquisition-phase and test-phase performance, but it is limited to the visual-spatial condition (and possibly to the auditory one as well).

Regression analyses

The results presented here suggest *quantitative* learning differences between sense modalities, which to some extent may be at least partly due to differences in perceiving and/or remembering sequences in each input condition. In order to detect more subtle modality-specific learning differences, we explored whether there also exist learning-related *qualitative* modality effects. Specifically, we conducted regression analyses to determine which sources of information may have been extracted in each modality/format condition. Each of the legal and illegal test items were assessed in terms of their initial and final anchor strengths (IAS and FAS), an indication of the relative frequencies of the initial and final fragment "chunks" (i.e., bi- and trigrams) that exist in similar positions in the training items (Conway & Christiansen, 2005). For example, the test item 1-2-1-3-5-2 has an IAS of 4.5 and an FAS of 2.0, indicating that the initial chunks 1-2, 2-1, and 1-2-1 occur frequently in the initial positions of the training set, whereas the final chunks 3-5, 5-2, and 3-5-2 occur slightly less frequently in the final positions of the training set.[2] We used SPSS 16.0 to run six different linear regression analyses, one for each modality/format condition. For each analysis, the dependent variable was number of endorsements for each test item, summed over all subjects. IAS, FAS, and item length were the three independent variables (predictors). A stepwise variable selection method was used (probability of entry = .05; probability of removal = 0.1). The results of the linear regression analyses indicate which of these three measures best predicts whether a participant in each modality/format condition will endorse a test item as legal. The results revealed a striking difference between the auditory and the visual conditions: FAS was a significant predictor for auditory endorsements at both presentation rates, $R^2 = .70$, $p < .001$ for slow and $R^2 = .64$, $p < .05$ for fast, whereas FAS was not a significant predictor for any of the visual conditions. On the other hand, IAS was a significant predictor for visual-spatial, $R^2 = .42$, $p < .01$, and visual-temporal endorsements, $R^2 = .36$, $p < .01$, at the slow rate of presentation. No other predictors were statistically significant for the slow visual conditions and none of the predictors were significant for the fast visual conditions. Finally, length was not a significant predictor for any of the conditions. These results therefore suggest that auditory statistical learning may rely most heavily on the fragment information contained at the endings of input sequences (i.e., a "recency" effect) whereas visual statistical learning is most sensitive to fragment information contained at the beginnings of input sequences (i.e., a "primacy" effect). Note that these qualitative

[2] In Conway and Christiansen (2005), we assessed these same stimuli in terms of additional information sources, such as *novelty*, *novel fragment position*, and *similarity*, but found that for this particular test and training set, these measures were highly correlated with IAS and FAS. A principle components analysis revealed that all these sources could be efficiently reduced to IAS, FAS, and item length. Therefore, we include only those three measures in the current regression analyses.

modality differences cannot be readily explained by differences in performance on the acquisition task.

GENERAL DISCUSSION

In summary, for the experimental group test data, there was a main effect of modality/format, with auditory performance highest and visual-temporal performance worst. We also found that at the fast presentation rate, only the visual-temporal condition was adversely affected. These quantitative learning differences were accompanied by qualitative learning effects. Consistent with previous results (Conway & Christiansen, 2005), audition and vision were differentially biased to encode statistical regularities towards the end and beginning of input sequences, respectively. There was some evidence that the quantitative modality effect was due at least in part to differences in the ability to adequately perceive and/or remember the input sequences. That is, perhaps auditory sequences and visual-spatial patterns are more easily perceived and held in memory, and thus the statistical regularities are easier to learn, relative to the visual-temporal input. However, this account cannot completely explain our pattern of results. At the fast presentation rate, the visual-temporal sequences were still readily perceived and remembered (as indicated by the acquisition phase task results), whereas the statistical regularities within the sequences were not learned (as indicated by the test-phase task results).

Overall, these results suggest that statistical learning is constrained by factors related to presentation modality, rate, and format (spatially vs. temporally distributed input). Participants in the visual conditions appeared to extract statistical patterns best when the input was presented in a spatial format rather than a temporal one. Additionally, visual learning relied upon statistical information present at the beginning of input sequences. In contrast, the auditory modality was quite adept at encoding temporal input, which was mediated by greater sensitivity to the statistical structure at the end of input sequences. Furthermore, the quantitative modality constraints were most pronounced at the fastest presentation rate, with visual-temporal learning adversely affected, highlighting the fact that vision is poor at encoding temporal regularities, at least at fast presentation rates.

The modality constraints observed here extend the results of Conway and Christiansen (2005), who found that auditory statistical learning of sequential structure exceeded that of visual or tactile learning. The results are also consistent with Saffran (2002), who found that learning predictive relationships occurred better for visual-simultaneous and auditory input compared to visual-sequential material. In the current study, visual learning matched auditory learning only when the presentation rate was slow.

Increasing the rate of presentation caused a decline in learning performance for the visual-temporal condition, whereas the auditory and visual-spatial groups showed no decline. These results illustrate that the extraction of statistical patterns is affected by the modality and presentation format in which the input is delivered. Furthermore, given the visual-temporal acquisition and test phase results, it appears that presentation rate has an effect on the *learning* of statistical patterns, over and above any effects it has on the *perceiving* or *remembering* of those same patterns. In and of itself, this finding is noteworthy because it suggests a relative independence of statistical learning from other processes related to general perception and memory.[3]

The differences in the initial/final sensitivities by vision and audition were also observed previously (Conway & Christiansen, 2005). It is important to note that for the current stimulus set, both fragment-initial and fragment-final information are equally helpful for making correct classification judgements; thus, it is unlikely that the observed qualitative sensitivity biases are due to experiment-specific learning. Instead, it is more likely that learners come into the experiment with preexisting auditory-final and visual-initial encoding biases. Interestingly, similar modality initial/final effects are seen in the realm of serial recall (i.e., serial position effects) with audition showing better recall for the end of lists and under certain circumstances, vision showing better recall for the beginning of lists (Beaman, 2002). Thus, there may exist a global perceptual/cognitive constraint on vision and audition, evolved and developed through currently unknown selection pressures, to differentially attend to information in initial and final sequence positions, respectively.

Taken together, these findings suggest that a full understanding of statistical learning and, it could be argued, other forms of perception, learning, and memory, will only occur by taking into account the effects of input modality in relation to the dimensions of space and time. The empirical evidence supports a *modality-constrained view* of cognition. That is, as previous work in attention, perception, learning, memory, and cognition has shown, various aspects of cognitive functioning appear to be constrained by factors having to do with the presentation modality of the input, which in turn may implicate the involvement of modality-specific, sensorimotor mechanisms and/or representations (cf. Barsalou, 1999; Collier & Logan, 2000; Conway & Christiansen, 2005, 2006; Duncan, Martens, & Ward, 1997; Freides, 1974; Gardiner & Cowan, 2003; Geldard, 1970; Glenberg, 1997; Harris, Petersen, & Diamond, 2001; Mahar et al.,

[3] It should be noted that the exact role that perception and short-term or working memory play in implicit statistical learning has not been extensively explored and thus remains a ripe avenue for future research.

1994; Mauk & Buonomano, 2004; Metcalfe, Glavanov, & Murdock, 1981; O'Connor & Hermelin, 1978; Penney, 1989; Repp & Penel, 2002; Rubin, 2006; Wilson, 2001). For the case of implicit statistical learning specifically, the current findings show that the ability to extract statistical regularities is strongly affected by the combination of presentation modality, presentation rate, and input format (i.e., spatial or temporal) of the stimuli in question.

Before concluding, we will address two possible concerns with the present results and interpretations. The first is of an empirical nature, the second, theoretical. The first concern is that the current empirical design is limited in that it does not fully compare learning across the various sense modalities in all input formats. For instance, there was not an auditory-spatial condition. Also, other types of visual or auditory stimuli were not included, such as abstract shapes or spoken nonsense syllables. Clearly, these are additional manipulations that are needed in order to fully investigate the role of space and time in visual and auditory statistical learning; we anticipate that the current work will provide the springboard for these future investigations. To our knowledge, aside from Conway and Christiansen (2005) and Saffran (2002), no other empirical study of AGL or statistical learning has made a rigorously controlled attempt to compare learning in different sense modalities and input formats as has been done here.[4]

The second concern regards the interpretation of vision and audition as being biased towards spatial and temporal patterns, respectively. In that vein, we should point out that although the modality effects shown here and elsewhere appear to be substantial and robust, there also exist commonalities between visual and auditory processing (e.g., Kubovy & van Valkenburg, 2001; Marks, 1978). Both vision and audition can be used to localise stimuli in space, to detect movement, to perceive rhythms and sequential patterns, and to discriminate objects based on when they occurred. As suggested from an ecological approach to perception (Gibson, 1966; Stoffregen & Bardy, 2001), such environmental information might be "amodal" in the sense that these and other features can be picked up across multiple sense modalities. If so, then different stimulus energies could be considered to be equivalent, or invariant, to one another. Similarly, Marks (1978) argued that the sense modalities have much in common in terms of phenomenological attributes, principles, and mechanisms. As he suggested, it is likely that the auditory and visual systems rely upon many of the same computational algorithms and neural architectures (cf. Shamma, 2001). From these perspectives,

[4] Pothos and Bailey (2001) and Pothos et al. (2006) did compare visual implicit learning using several different types of presentation formats and found no effects on learning. However, all formats consisted of visual regularities that were spatially distributed; there were no conditions examining temporally distributed regularities or auditory input.

perhaps the vision–space, audition–time analogy is misleading (Handel, 1988).

Clearly, it may be misleading to consider vision solely within a spatial framework and audition as existing only within a temporal one. However, there appear to be real biases in terms of which dimensions are more or less important for each modality. From a purely phenomenological perspective, it seems nearly impossible to imagine an atemporal sound or a nonspatial visual percept. Time appears to be the primary foundation for audition, with sounds changing in certain ways over time, whereas space is the primary referent for vision, with visual objects defined by size and shape (Hirsh, 1967).

Thus, however much the sensory systems may be similar, it is clear that certain constraints affect their processing, and therefore it is important to understand the nature of the constraints. Within the context of verbal short-term memory, Penney (1989) suggested that auditory and visual stimuli are processed in separate "streams" with each having different properties and capabilities. The auditory stream is characterised by strong associations between successive/sequential items, whereas the visual stream most readily encodes associations between simultaneously presented spatial input. This makes sense in light of what is known about the principles of auditory grouping: Sounds that rapidly follow one another tend to be produced by the same environmental source and thus are perceived together, whereas it is spatial rather than temporal contiguity which is important for visual perception. Possibly, a similar associative neural learning mechanism may be operative in both auditory and visual cortical areas, but each makes its computations over a different dimension of input. The existence of similar underlying computational principles helps explain the processing similarities across the modalities, while the proposed separate processing streams allow for the existence of modality-specific differences. Within this framework, visual mechanisms might more readily associate features and/or objects that are close together in space, whereas auditory mechanisms would encode relations among events occurring close together in time. The origin of these modality constraints is currently unknown and unexplored. Presumably, they are at least partly experience dependent, developing over a lifetime of exposure to various sensory stimuli occurring in different presentation formats, with the end result being a suite of learning systems that have each become tuned to focus on particular aspects of environmental stimuli over others.

As has been argued elsewhere (Barsalou, 1999; Conway & Christiansen, 2006; Rubin, 2006), any theory of cognition that cannot account for modality-specific phenomena will not be adequate. As the field of cognitive science progressively moves away from the "mind-as-amodal-and-disembodied-computer" metaphor, it becomes increasingly apparent that understanding the role played by the sensory systems is necessary to obtain a full theory of human memory and perception. Although additional work is needed to investigate implicit statistical learning in all modality/format combinations (e.g., auditory-spatial input), the current data shed light on one important learning constraint: Statistical learning is heavily affected by the sense modality and presentation format of the stimuli in question, especially at fast presentation rates. The presence of this constraint suggests that auditory and visual statistical learning may at least partly depend upon modality-specific processing streams, each having different biases and properties. Additional research must further elucidate the nature of each modality-constrained learning system and how they support human cognition more broadly.

Original manuscript received July 2007
Manuscript accepted March 2008
First published online July 2008

REFERENCES

Altmann, G. T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences*, *99*, 15250–15251.

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.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology*, *55A*, 371–389.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General*, *120*, 316–323.

Chang, G. Y., & Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 714–722.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416.

Cleeremans, A., & McClelland, J. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235–253.

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory and Cognition*, *28*, 529–538.

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.

Conway, C. M., Goldstone, R. L., & Christiansen, M. H. (2007). Spatial constraints on visual statistical learning of multi-element scenes. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 185–190). Austin, TX: Cognitive Science Society.

Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, *12*, 317–334.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 756–757.

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

Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, *81*, 284–310.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition*, *22*, 95–110.

Gardiner, J. M., & Cowan, N. (2003). Modality effects. In J. H. Byrne, H. Eichenbaum, H. L. Roediger III, & R. F. Thompson (Eds.), *Learning and memory* (pp. 397–400). New York: Macmillan.

Geldard, F. A. (1970). Vision, audition, and beyond. In W. D. Neff (Ed.), *Contributions to sensory physiology* (Vol. 4, pp. 1–17). New York: Academic Press.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*, 278–285.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

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

Glenberg, A. M., & Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 3–15.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology*, *33*, 154–207.

Handel, S. (1988). Space is to time as vision is to audition: Seductive but misleading. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 315–317.

Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology*, *21*, 256–266.

Harris, J. A., Petersen, R. S., & Diamond, M. E. (2001). The cortical distribution of sensory memories. *Neuron*, *30*, 315–318.

Hirsh, I. J. (1967). Information processing in input channels for speech and language: The significance of serial order of stimuli. In F. L. Darley (Ed.), *Brain mechanisms underlying speech and language* (pp. 21–38). New York: Grune & Stratton.

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., & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

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

Marks, L. E. (1978). *The unity of the senses: Interrelations among the modalities.* New York: Academic Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J.-R., & Druhan, B. (1989). The role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1083–1100.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, *27*, 307–340.

Metcalfe, J., Glavanov, D., & Murdock, M. (1981). Spatial and temporal processing in the auditory and visual modalities. *Memory and Cognition*, *9*, 351–359.

Meulemans, T., & van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1007–1028.

O'Connor, N., & Hermelin, B. (1978). *Seeing and hearing and space and time*. New York: Academic Press.

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

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General*, *130*, 401–426.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and 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.

Pothos, E. M., & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 847–862.

Pothos, E. M., Chater, N., & Ziori, E. (2006). Does stimulus appearance affect learning? *American Journal of Psychology*, *119*, 277–301.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Behavior*, *6*, 855–863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, *125*, 123–138.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1085–1099.

Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science*, *1*, 277–311.

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., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, *5*, 340–348.

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

Stadler, M. A., & Frensch, P. A. (Eds.). (1998). *The handbook of implicit learning*. London: Sage.

Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*, 195–261.

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence of the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1322–1333.

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.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review*, *8*, 44–57.

APPENDIX A

Acquisition pairs

Matched pairs	*Mismatched pairs*
1-2-1-1-1-3/1-2-1-1-1-3	1-2-3-5-2-5/1-2-3-5-2-3
4-1-1-3-5-2/4-1-1-3-5-2	1-2-3-5-2-3/1-2-3-5-2-5
4-1-3-5-2/4-1-3-5-2	4-3-5-2-3/4-3-5-2-5
1-2-5-5-5/1-2-5-5-5	4-3-5-2-5/4-3-5-2-3
4-1-3/4-1-3	1-2-5-5/1-2-1-3
1-2-3/1-2-3	1-2-1-3/1-2-5-5

The numbers refer to a particular visual or auditory stimulus (see text).

APPENDIX B

Test items

Legal	*IAS*	*FAS*	*Illegal*	*IAS*	*FAS*
4-1-3-5-2-3	2.5	2.5	1-4-5-1-3-3	0	0
1-2-1-3-5-2	4.5	2.0	4-5-1-2-1-3	0	2.0
4-3-5-2-5-5	2.0	1.5	4-2-1-3-1-5	0	0
4-1-3-5-2-5	2.5	2.0	1-5-3-3-2-2	0	0
4-1-1-1-3	2.0	2.0	1-5-3-4-2	0	0
1-2-1-1-3	4.5	2.0	4-2-1-5-3	0	0
1-2-3-5-2	5.0	2.0	1-5-3-1-2	0	0
4-1-1-3	2.0	2.0	4-5-1-3	0	1.5
4-3-5-2	2.0	2.0	4-5-2-2	0	0
1-2-5	4.5	1.0	1-4-2	0	0

The numbers refer to a particular visual or auditory stimulus (see text). IAS = initial anchor strength; FAS = final anchor strength.