Perception advance online publication

doi:10.1068/p5843

Simultaneous and independent acquisition of multisensory and unisensory associations

Aaron R Seitz, Robyn Kim¶, Virginie van Wassenhove¶§, Ladan Shams¶
Department of Psychology, Boston University, 64 Cummington Street, Boston, MA 02215, USA;
e-mail: aseitz@bu.edu; ¶ Department of Psychology, University of California at Los Angeles,
Los Angeles, CA 90095, USA; § Division of Biology, California Institute of Technology, Pasadena,
CA 91125, USA
Received 29 September 2006, in revised form 18 April 2007; published online 12 September 2007

Abstract. Although humans are almost constantly exposed to stimuli from multiple sensory modalities during daily life, the processes by which we learn to integrate information from multiple senses to acquire knowledge of multisensory objects are not well understood. Here, we present results of a novel audio – visual statistical learning procedure where participants are passively exposed to a rapid serial presentation of arbitrary audio – visual pairings (comprised of artificial/synthetic audio and visual stimuli). Following this exposure, participants were tested with a two-interval forced-choice procedure in which their degree of familiarity with the experienced audio – visual pairings was evaluated against novel audio – visual combinations drawn from the same stimulus set. Our results show that subjects acquire knowledge of visual – visual, audio – audio, and audio – visual stimulus associations and that the learning of these types of associations occurs in an independent manner.

1 Introduction

Our sensory systems are constantly bombarded with a complex array of information pertaining to our physical environment. Somehow, we are able to make sense of these inputs and structure our behavior accordingly. This feat is particularly impressive, given the vast amount of information that we are able to encode. For instance, estimates of human memory capacity range from 10^9 (Laudauer 1986) to 10^{20} (Von Neumann 1958) bits, which translates to hundreds of thousands of visual objects, tens of thousands of words and other sounds, and many skills, episodic memories, etc. How we acquire this knowledge, and what factors constrain learning, is a matter of intensive study (for reviews see Chun 2000; Das et al 2001; Goldstone 1998; Seitz and Dinse 2007; Seitz and Watanabe 2005; Wasserman and Miller 1997).

Recent lines of research demonstrate conditions in which learning is patterned after the statistical regularities of the surrounding environment in the absence of any explicit task (Dinse et al 2003; Frenkel et al 2006; Saffran et al 1999). For instance, theoretical studies of early development suggest that the patterning of the early sensory systems can be accounted for by the statistical regularities of the sensory environment (ie by the pattern of inputs received by the system) and that, in turn, the distribution of feature detectors in the sensory structures is determined by the frequency distributions of such features in the physical environment (Grossberg 1976; Grossberg and Seitz 2003; Kohonen 1982; von der Malsburg 1973). In accordance with this viewpoint, animals that are raised in environments with restricted visual experience (Blasdel et al 1977; Hubel and Wiesel 1964; Singer et al 1981), or restricted auditory experience (Chang et al 2005; King and Moore 1991; King et al 2000) during their developmental critical periods develop biased representations in primary sensory cortical areas.

While these developmental studies present an extreme case of how sensory alterations affect sensory learning, studies of 'statistical learning' (Fiser and Aslin 2001; Saffran et al 1999) examine how stimulus – stimulus associations can be rapidly learned solely on the basis of inter-stimulus statistical contingencies. In this paradigm, participants are passively exposed to arbitrary stimulus configurations for which the underlying statistical properties are manipulated by the experimenter. After passive exposure to the stimuli, participants' degree of familiarity with the experienced stimulus combinations is evaluated. These studies have demonstrated that a relatively brief exposure to a set of visual (Fiser and Aslin 2001, 2002) or auditory (Saffran et al 1999) stimulus features allows subjects to learn statistical relationships among the stimuli. For instance, stimuli that are consistently presented in spatial (Fiser and Aslin 2001) or temporal (Fiser and Aslin 2002) proximity become more familiar than those that are not correlated in space or time. While statistical learning has been shown for visual (Fiser and Aslin 2001), auditory (Saffran et al 1999), and somatosensory (Conway and Christiansen 2005) modalities, cross-modal statistical learning has not yet been investigated.

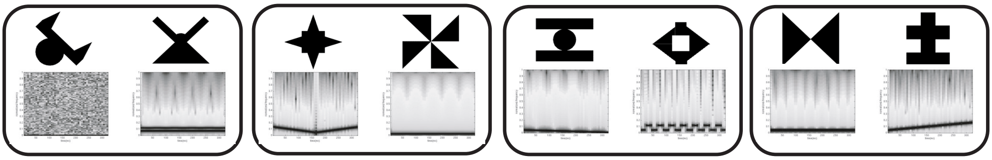
In this study, we address four key issues related to audio–visual statistical learning. First, does the learning of associations between stimuli in the same sensory modality occur when information is available concurrently in more than one sensory modality? Second, can audio–visual associations be learned solely on the basis of the statistics of stimulus co-occurrence? Third, is the degree of learning a function of the temporal rate in the audio and visual streams? Fourth, to what degree is the learning of unisensory associations affected by the simultaneous learning of other modal or crossmodal associations?

2 Experiment 1

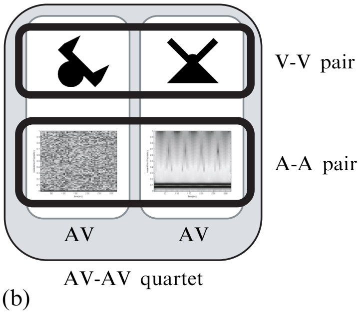
2.1 *Methods*

2.1.1 *Participants.* Thirty-six undergraduate participants (aged 18–35 years) (nine in each group) took part in the study and received course credit or payment for their participation. All participants were naive to the purpose of the study and participated in only one experimental condition. The experiments were conducted in accordance with the IRB approved by the Committee on Human Research of the University of California and the Declaration of Helsinki.

2.1.2 *Stimuli.* Visual stimuli consisted of a set of eight arbitrary black-and-white figures (shown in top row of figure 1a) adapted from Fiser and Aslin (2001). Images were sized to 128 × 128 pixels, subtended 3 deg, and were each presented for 300 ms. Visual stimuli were presented on a 19-inch CRT monitor with resolution of 1152 × 768 and refresh rate of 75 Hz.



(a)



(b)

Figure 1. Sample of stimuli used in experiments 1 and 2. Eight shapes and eight sounds were used in experiments 1 and 2. (a) Top row: all eight stimuli, from each modality, are shown in example ensembles. The actual subcomponents of the quartets were randomly assigned for each participant. Bottom row: the spectrograms (abscissa is time, ordinate is normalized frequency) of the audio stimuli. (b) For illustration purpose, one stimulus-ensemble is shown with designations for each of the three pairs (A-A, V-V, AV) and the quartet.

The 8 audio stimuli with differing spectrotemporal properties created in Matlab (v. 7, Mac OS X) were designed to mimic the properties of the visual stimuli—easily discriminable, and unfamiliar. Spectrograms of all stimuli used are shown in the bottom row of figure 1a. The duration of the audio stimuli was 300 ms (5 ms on/off ramp). Audio stimuli were presented binaurally via speakers positioned on each side of the monitor screen at the same height as the visual stimuli. Sound pressure level was set at a comfortable hearing level of ~60–65 dB for all participants.

Audio and visual stimuli were presented simultaneously during the first part of the experiment ('exposure', see below), with each of the 8 audio and 8 visual stimuli uniquely assigned to a specific 'quartet' (a sequence of 2 audio–visual synchronous pairs; see figure 1b for details). In other words, stimuli were presented as a stream of audio–visual quartets (AV-AV), which consisted of visual pairs of sequentially presented shapes (V-V) that always co-occurred with a certain pair of sequentially presented sounds (A-A). The stimulus make-up for these quartets was randomly assigned for each subject. During the second part of the experiment ('testing', see below), test stimuli consisted of either audio–visual quartets (AV-AV), unisensory sequential pairs (V-V or A-A), or synchronous audio–visual pairs (AV).

2.1.3 *Procedure.* The experiments took place in a dimly lighted room. Participants sat 57 cm away from the monitor screen with their heads stabilized with a chin-rest. Stimuli were presented with custom software written with use of the Psychophysics Toolbox (Brainard 1997; Pelli 1997) for Matlab™ (Natick, MA) on a Macintosh G4 computer.

During the first phase of the experiment ('exposure', ~8 min long), participants were presented with a rapid serial presentation of a continuous stream of four AV-AV quartets presented one hundred times each in a pseudorandom order with the constraint that a given quartet could not appear twice in immediate succession. Quartets or pairs could not be segmented on the basis of any temporal or spatial cues as the SOA of the AV stimuli was fixed within and across quartets. Four different inter-stimulus-interval (ISI) conditions (0, 100, 300, and 500 ms) were run in different subject groups. Each subject was randomly assigned to an ISI condition, which specified the ISI between each AV pair (and, accordingly, between sequential quartets). The same ISI was used both for exposure and for testing. Subjects were asked to carefully watch and listen to the stimuli. At the time of exposure, participants were not informed that there would be a subsequent testing phase.

Following exposure, in the testing phase, participants were tested with a two-interval forced-choice (2IFC) paradigm, judging in each trial which of the two intervals contained the stimulus that was most familiar to them. During the testing phase, the stimuli were either presented as unimodal pairs (A-A or V-V), bimodal pairs (AV), or bimodal quartets (AV-AV). In a given trial, one interval consisted of one ensemble (A-A, V-V, AV, or AV-AV) that had been repeatedly presented during exposure, and another interval consisted of an ensemble of the same sensory combination (A-A, V-V, AV, or AV-AV, respectively) constructed of a novel combination of shape and/or sounds randomly selected from the stimulus set used during exposure (in some cases the same stimulus element could be used in both the novel and the exposed ensemble). Specifically, while during exposure the joint probability of the presentation of each exposed AV-AV quartet (or each A-A or V-V pair) was 0.125 (and a conditional probability of constituent elements equal to 1), the joint (and conditional) probability of the novel ensembles was 0. Participants were allowed 2 s to provide their choices (1 or 2) by key press. 8 repetitions of each of the 4 stimulus conditions for each ensemble were presented in a randomly interleaved fashion for a total of 128 trials per subject.

2.2 *Results and discussion*

2.2.1 *Joint audio – visual probabilities.* The goal of the first experiment was to determine whether individuals are sensitive to the conjunction of audio – visual (AV) events, ie to their joint probability. During exposure to the audio – visual stimuli, the relative timing of the audio and visual streams was arranged such that A-A pairs always co-occurred with the V-V pairs, constituting audio – visual quartets (AV-AV). If the two streams were presented separately, previous work would predict that observers would learn visual associations (V-V pairs) (Fiser and Aslin 2001) or, similarly, audio associations (A-A pairs) (Saffran et al 1999). Here, however, two streams were presented concurrently. Because this paradigm could disrupt learning (Seitz et al 2005) within the individual sensory modalities (for instance owing to attentional overload or interference), we also checked whether unisensory associative learning (A-A and V-V) could occur simultaneously with multisensory associative learning (AV).

The first question that we addressed was whether learning of unimodal pairs occurs when information is available concurrently in more than one modality. Figure 2a shows the average performance of all participants collapsed across the different ISI conditions. A significant learning effect was observed for the A-A pairs ($t_{35} = 3.7$, $p < 0.001$, two-tailed t-test versus chance level of 50%) and for the V-V pairs ($t_{35} = 3.9$, $p < 0.001$, t-test). This suggests that A-A and V-V associations can develop concurrently through statistical learning.

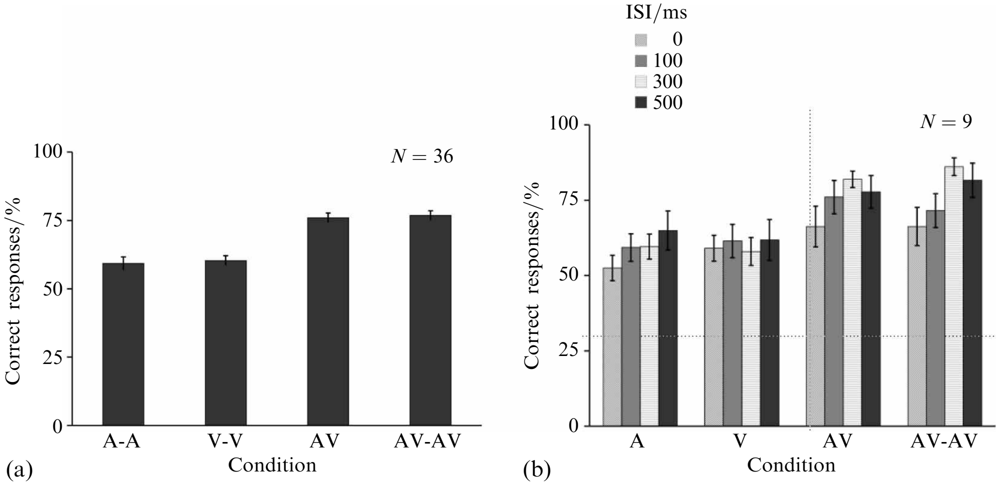


Figure 2. Experiment 1: learning of audio, visual, and audio – visual ensembles. (a) Mean performance across all subjects for audio, visual, audio – visual pairs, and audio – visual quartets (averaged across ISI conditions). (b) Effect of ISI: mean performance across all subjects for audio, visual, audio – visual pairs, and audio – visual quartets. Error bars indicate 1 SEM.

Our second question was whether audio – visual ensembles can be learned. Collapsing data across the ISI conditions showed that there was a significant learning effect for the AV pairs ($t_{35} = 9.5$, $p < 0.001$, t-test). These results provide a first demonstration that statistical learning can indeed produce multisensory associations. We found a greater degree of learning for the crossmodal pairs (AV) when compared to the unisensory pairs: AV versus A-A ($t_{35} = 6.3$, $p < 0.001$, t-test, paired), AV versus V-V ($t_{35} = 5.3$, $p < 0.001$).

Additionally, the performance for AV-AV quartets was significantly better than chance ($t_{35} = 9.3$, $p < 0.001$, two-tailed t-test) and better than the performance in the unisensory conditions: AV-AV versus A-A ($t_{35} = 5.5$, $p < 0.001$, t-test, unpaired) and AV-AV versus V-V ($t_{35} = 4.9$, $p < 0.001$). An important consideration in evaluating the performance of the AV-AV quartets is that subjects could be responding to the quartets by separately processing the A-A and the V-V pairs (ie responding to the subcomponents of the quartets). If this were the case, performance on the AV-AV quartets should be less than, or equal to, the summed probability of responding correctly to each of the unimodal pairs. Indeed, subjects did perform slightly worse ($t_{70} = -1.7$, $p < 0.05$, paired t-test) on the AV-AV quartets (mean, M = 77%; standard error, SE = 5.7%) than the summed probability of the A-A and V-V pairs (M = 84%, SE = 9.4%). The performance on the quartets also did not significantly differ from performance on the AV pairs ($t_{35} = 0.47$, $p = 0.64$, paired t-test). Nonetheless, participants' high performance on the AV pairs, which do not contain unisensory pairs (A-A or V-V) as subcomponents, cannot be explained by summing the probabilities on the unimodal pairs, and thus unequivocally reflects the learning of multisensory associations.

It is also notable that our testing sessions are rather long, and with 8 repetitions per pair (or 32 trials for each condition), there exists the concern that some part of the learning could occur during the testing sessions. To test for this possibility, we examined whether performance differed between first and second half of the testing sessions. Collapsing across all ISI conditions, we found no difference between session parts in the AV-AV, V-V, and AV conditions ($t_{35} = 0.8231$, 1.0428, 1.4620; $p = 0.42$, 0.30, 0.15, paired t-tests, respectively). In the case of the A-A condition a significant effect was observed ($t_{35} = 2.2$, $p < 0.05$); however, performance here actually decreased in the second half (first half, M = 62%, SE = 3%; second half, M = 57%, SE = 3%). This deterioration of learning is in the opposite direction of what would be expected if the testing session contributed to the learning effect.

Our third question addressed the extent to which the rate of stimulus presentation affects the degree of statistical learning. Overall, participants performed better as the ISI increased and the pattern of performance across ensemble types did not change as a function of ISI (see figure 2b). This observation was substantiated with a 2-way ANOVA showing a significant main effect of ISI ($F_{3,143} = 12.95$, $p < 0.001$) and of ensemble type ($F_{3,143} = 2.7$, $p < 0.05$). However, the interaction of ISI and ensemble type yielded no significant effect ($F_{9,143} = 0.65$, $p = 0.75$). Thus, although ISI affected learning, it did not have a significant impact on the type of ensembles that were best learned.

Our fourth and final question regarded the degree to which the learning of unisensory audio or visual associations is impacted by the simultaneous acquisition of multisensory associations and unisensory associations of the other modality. We therefore did a second experiment to address these issues.

3 Experiment 2

3.1 *Methods*

3.1.1 *Participants.* Forty-eight new undergraduate participants (aged 18 to 35 years) (twelve in each group) took part in the study and were treated in the same manner as described in section 2.1.1.

3.1.2 *Stimuli.* The same stimuli were used here as described in section 2.1.2.

3.1.3 *Procedure.* Two multisensory groups were exposed to the identical exposure procedure as in the first experiment (with 300 ms ISI). However, one group was tested only on visual pairs (AV-vistest group), while the other group was tested only on audio pairs (AV-audtest group). The remaining two groups were run in the unisensory condition, where one group of subjects was exposed to only the visual stream and tested on visual pairs (V-vistest) and the other group was exposed to only the audio stream and was tested on audio pairs (A-audtest). One subject in the V-vistest group was dropped from the study for failure to respond within the 2 s response window of the test trials and thus only data from eleven subjects were analyzed from the V-vistest group. During testing, 8 repetitions per unisensory pair (visual or audio) were presented for a total of 32 trials per subject.

3.2 *Results and discussion*

3.2.1 *Impact of multisensory learning on unisensory learning.* The results of experiment 2 are shown in figure 3. We found a remarkable similarity between the performance of subjects in the multisensory and unisensory exposure groups. Subjects in the audio-test groups performed on average 66% ± 4% for the AV-audtest and 66% ± 5% for the A-audtest, with no significant difference between the groups ($t_{22} = 0.1$, $p = 0.92$, t-test). Similarly, subjects in the visual-test groups performed on average 66% ± 3% for the AV-vistest and 65% ± 5% for the V-vistest, with no significant difference between the groups ($t_{21} = 0.08$, $p = 0.94$, t-test). These results are higher than those found in the first experiment (A-A = 60% ± 4%, V-V = 58% ± 5%), but, given that subjects in experiment 1 experienced interleaved trials from each of the four different types of associations, it seems likely that the performance differences between experiments are due to the differences between the testing procedures. The results of experiment 2 thus demonstrate that simultaneous learning of other-modal and crossmodal associations does not negatively impact the acquisition of unisensory visual or audio associations, and that these three types of associations can develop independently.

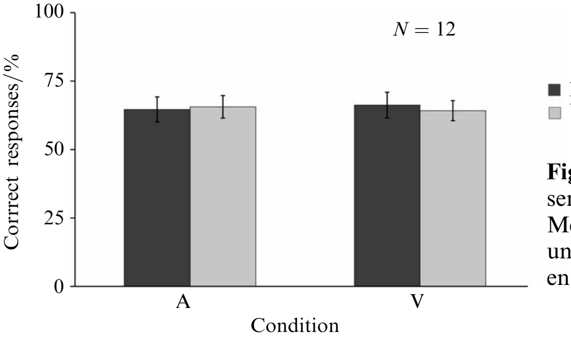


Figure 3. Experiment 2: Impact of multisensory learning on unisensory learning. Mean performance in the multisensory and unisensory groups for audio and visual ensembles. Error bars indicate 1 SEM.

4 General discussion

The different sensory modalities are generally held to be independent and self-contained. Recent research, however, demonstrates that crossmodal interactions are ubiquitous in human perception (Driver 1996; Driver and Spence 1998; Howard and Templeton 1966; McGurk and MacDonald 1976; Sekuler et al 1997; Shams et al 2000; Soto-Faraco et al 2002; Spence et al 2004; Violentyev et al 2005), and has led to a gradual shift from a paradigm of sensory modularity to a more interactive and integrative paradigm of multisensory processing (Calvert et al 2004; Ghazanfar and Schroeder 2006; Shimojo and Shams 2001). Despite the overwhelming evidence for crossmodal interactions in various tasks, multisensory and crossmodal learning has been by and large neglected (but see Goulet and Murray 2001; Lacey and Campbell 2006; Tanabe et al 2005).

Our results provide the first demonstration of multisensory statistical learning. In experiment 1, we showed that multisensory statistical learning of AV associations occurs in parallel with unisensory statistical learning of A-A and V-V associations and that the magnitude, but not the pattern, of learning is affected by the rate of presentation. We also found that observers discriminated the AV pairs significantly better than the unisensory pairs (A-A or V-V), but it remains unclear whether this superiority reflects a better propensity to learn crossmodal associations or reflects the finding that concurrently presented stimuli are better learned than successively paired stimuli.

Importantly, the simultaneous learning of other-modal and crossmodal associations did not negatively impact the acquisition of unisensory visual or audio associations. Experiment 1 demonstrated that learning of multisensory associations occurred simultaneously with learning of unisensory associations. Furthermore, experiment 2 verified that the degree of learning of unisensory associations was independent of multisensory learning, since unisensory learning was equivalent for subjects exposed to multisensory and unisensory associations.

Our findings are in line with a recent study showing evidence of independent acquisition of audio and visual artificial grammars through statistical learning (Conway and Christiansen 2006). While this study showed a similar pattern of independent acquisition of grammars in different modalities, it did not probe crossmodal sequences. In addition, our study differs in the fact that our audio and visual stimuli were presented in simultaneous streams, and hence we show that learning of AV as well as A-A and V-V associations can develop in parallel.

In addition to the demonstration of multisensory statistical learning, the current results provide some hints about the properties underlying this type of associative learning. For instance, our results are consistent with the existence of three similar but separate associational systems (one audio, one visual, and one multisensory). This postulate of separate associational systems for unisensory and multisensory events is consistent with recent research in attentional processing systems. For instance, while some studies show strong evidence of multisensory attentional systems (Busse et al 2005; Driver and Spence 1998; Spence and Driver 1996), other research demonstrates that attentional effects can also be modality specific (Beer and Roder 2005; Duncan et al 1997). Alternatively, associative learning could be subserved by a learning process that operates on representations in a modality-independent manner, and thus can result in learning of associations regardless of modality of origin. Such a multimodal or amodal learning mechanism would be consistent with the remarkable similarity in the degree of learning between the A-A and the V-V associations. However, since we found that learning in one modality did not affect learning in others, one would need to incorporate the assumption that associations within (and perhaps across) modalities are learned independently. Furthermore, to explain the different magnitude of learning between the unisensory and multisensory associations in experiment 1, this system would need to be more effective in extracting relationships among simultaneous stimuli compared to successive stimuli. Further research will be necessary to distinguish between the three-systems and single-system hypotheses.

Statistical learning is often described as a type of implicit learning (Fiser and Aslin 2001, 2002). To promote the implicitness of learning, we adopted the use of a rapid serial presentation, as the temporal masking and interference with memory make it difficult for subjects to consciously keep track of the stimulus ensembles. Similar rapid serial presentations have been used in other studies of statistical learning (Conway and Christiansen 2005; Saffran et al 1999). Indeed, we found, through informal interviews, that subjects tended to be unaware of their level of performance; even subjects who performed well above chance tended to report that they had low confidence in their responses. This leads us to believe that our results reflect an automatic, implicit, learning process rather than a conscious assessment of the stimuli. While ecological validity of psychophysical tasks is always questionable, we believe the paradigms employed in our experiments are not too far removed from natural settings, as similar learning processes are likely to operate in learning navigational cues or other regularities when rapidly traversing new environments (eg running through the woods).

5 Conclusion

Our experience with the world intrinsically involves the simultaneous processing of information from multiple sensory modalities. The present study demonstrates that associations, both within and across modalities, can be built implicitly on the basis of the statistics of stimulus presentation. Furthermore, multisensory and unisensory learning may occur in parallel, which suggests that statistical learning may operate on all stimuli regardless of their modality of origin. Taken together, our results are an important step in understanding how humans learn to combine information within and across sensory modalities to acquire new knowledge of their environment.

Acknowledgments. We thank Jozsef Fiser for kindly providing the visual stimuli used in his experiments to us. This project was partly supported by a Naval Research Laboratory grant (N00173-06-1-G016) to LS, and AS was supported by NIH grant (R01EY015980-01).

References

Beer A L, Roder B, 2005 "Attending to visual or auditory motion affects perception within and across modalities: an event-related potential study" *European Journal of Neuroscience* **21** 1116 – 1130

Blasdel G G, Mitchell D E, Muir D W, Pettigrew J D, 1977 "A physiological and behavioural study in cats of the effect of early visual experience with contours of a single orientation" *Journal of Physiology* **265** 615 – 636

Brainard D H, 1997 "The Psychophysics Toolbox" *Spatial Vision* **10** 433 – 436

Busse L, Roberts K C, Crist R E, Weissman D H, Woldorff M G, 2005 "The spread of attention across modalities and space in a multisensory object" *Proceedings of the National Academy of Sciences of the USA* **102** 18751 – 18756

Calvert G, Spence C, Stein B E, 2004 *The Handbook of Multisensory Processes* (Cambridge, MA: MIT Press)

Chang E F, Bao S, Imaizumi K, Schreiner C E, Merzenich M M, 2005 "Development of spectral and temporal response selectivity in the auditory cortex" *Proceedings of the National Academy of Sciences of the USA* **102** 16460 – 16465

Chun M M, 2000 "Contextual cueing of visual attention" *Trends in Cognitive Sciences* **4** 170 – 178

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

Das A, Franca J G, Gattass R, Kaas J H, Nicolelis M A, Timo-Iaria C, Vargas C D, Weinberger N M, Volchan E, 2001 "The brain decade in debate: VI. Sensory and motor maps: dynamics and plasticity" *Brazilian Journal of Medical and Biological Research* **34** 1497 – 1508

Dinse H R, Ragert P, Pleger B, Schwenkreis P, Tegenthoff M, 2003 "Pharmacological modulation of perceptual learning and associated cortical reorganization" *Science* **301** 91 – 94

Driver J, 1996 "Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading" *Nature* **381** 66 – 68

Driver J, Spence C, 1998 "Crossmodal attention" *Current Opinion in Neurobiology* **8** 245 – 253

Duncan J, Martens S, Ward R, 1997 "Restricted attentional capacity within but not between sensory modalities" *Nature* **387** 808 – 810

Fiser J, Aslin R N, 2001 "Unsupervised statistical learning of higher-order spatial structures from visual scenes" *Psychological Science* **12** 499 – 504

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

Frenkel M Y, Sawtell N B, Diogo A C, Yoon B, Neve R L, Bear M F, 2006 "Instructive effect of visual experience in mouse visual cortex" *Neuron* **51** 339 – 349

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

Goldstone R L, 1998 "Perceptual learning" *Annual Review of Psychology* **49** 585 – 612

Goulet S, Murray E A, 2001 "Neural substrates of crossmodal association memory in monkeys: the amygdala versus the anterior rhinal cortex" *Behavioral Neuroscience* **115** 271 – 284

Grossberg S, 1976 "On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems" *Biological Cybernetics* **21** 145 – 159

Grossberg S, Seitz A, 2003 "Laminar development of receptive fields, maps and columns in visual cortex: the coordinating role of the subplate" *Cerebral Cortex* **13** 852 – 863

Howard I P, Templeton W B, 1966 *Human Spatial Orientation* (London: Wiley and Sons)

Hubel D H, Wiesel T N, 1964 "Effects of monocular deprivation in kittens" *Naunyn-Schmeidebergs Archiv für Experimentelle Pathologie und Pharmacologie* **248** 492 – 497

King A J, Moore D R, 1991 "Plasticity of auditory maps in the brain" *Trends in Neuroscience* **14** 31 – 37

King A J, Parsons C H, Moore D R, 2000 "Plasticity in the neural coding of auditory space in the mammalian brain" *Proceedings of the National Academy of Sciences of the USA* **97** 11821 – 11828

Kohonen T, 1982 "Self-organized formation of topologically correct feature maps" *Biological Cybernetics* **43** 59 – 69

Lacey S, Campbell C, 2006 "Mental representation in visual/haptic crossmodal memory: evidence from interference effects" *Quarterly Journal of Experimental Psychology* **59** 361 – 376

Laudauer T K, 1986 "How much do people remember? Some estimates of the quantity of learned information in long-term memory" *Cognitive Science* **10** 477 – 493

McGurk H, MacDonald J W, 1976 "Hearing lips and seeing voices" *Nature* **264** 746 – 748

Malsburg C von der, 1973 "Self-organization of orientation sensitive cells in the striate cortex" *Kybernetik* **14** 85 – 100

Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: transforming numbers into movies" *Spatial Vision* **10** 437 – 442

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

Seitz A R, Dinse H R, 2007 "A common framework for perceptual learning" *Current Opinion in Neurobiology* **17** 148 – 153

Seitz A, Watanabe T, 2005 "A unified model for perceptual learning" *Trends in Cognitive Sciences* **9** 329 – 334

Seitz A R, Yamagishi N, Werner B, Goda N, Kawato M, Watanabe T, 2005 "Task-specific disruption of perceptual learning" *Proceedings of the National Academy of Sciences of the USA* **102** 14895 – 14900

Sekuler R, Sekuler A B, Lau R, 1997 "Sound alters visual motion perception" *Nature* **385** 308

Shams L, Kamitani Y, Shimojo S, 2000 "What you see is what you hear" *Nature* **408** 788

Shimojo S, Shams L, 2001 "Sensory modalities are not separate modalities: plasticity and interactions" *Current Opinion in Neurobiology* **11** 505 – 509

Singer W, Freeman B, Rauschecker J, 1981 "Restriction of visual experience to a single orientation affects the organization of orientation columns in cat visual cortex. A study with deoxyglucose" *Experimental Brain Research* **41** 199 – 215

Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A, 2002 "The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities" *Cognitive Brain Research* **14** 139 – 146

Spence C, Driver J, 1996 "Audiovisual links in endogenous covert spatial attention" *Journal of Experimental Psychology: Human Perception and Performance* **22** 1005 – 1030

Spence C, Pavani F, Maravita A, Holmes N, 2004 "Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: evidence from the crossmodal congruency task" *Journal of Physiology (Paris)* **98** 171 – 189

Tanabe H C, Honda M, Sadato N, 2005 "Functionally segregated neural substrates for arbitrary audiovisual paired-association learning" *Journal of Neuroscience* **25** 6409 – 6418

Violentyev A, Shimojo S, Shams L, 2005 "Touch-induced visual illusion" *NeuroReport* **16** 1107 – 1110

Von Neumann J, 1958 *The Computer and the Brain* (New Haven, CT: Yale University Press)

Wasserman E A, Miller R, R, 1997 "What's elementary about associative learning" *Annual Review of Psychology* **48** 573 – 607

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

www.perceptionweb.com