

---

# Cross-modal interaction between vision and touch: the role of synesthetic correspondence

---

Gail Martino<sup>¶</sup>, Lawrence E Marks<sup>#</sup>

The John B Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519, USA;  
e-mail: [gmartino@jbpierce.org](mailto:gmartino@jbpierce.org); <sup>¶</sup> also Department of Diagnostic Radiology, Yale University School of Medicine and Department of Neuroscience, University of Connecticut Health Center; <sup>#</sup> also Department of Epidemiology and Public Health and Department of Psychology, Yale University  
Received 22 September 1999, in revised form 14 February 2000

---

**Abstract.** At each moment, we experience a melange of information arriving at several senses, and often we focus on inputs from one modality and ‘reject’ inputs from another. Does input from a rejected sensory modality modulate one’s ability to make decisions about information from a selected one? When the modalities are vision and hearing, the answer is “yes”, suggesting that vision and hearing interact. In the present study, we asked whether similar interactions characterize vision and touch. As with vision and hearing, results obtained in a selective attention task show cross-modal interactions between vision and touch that depend on the synesthetic relationship between the stimulus combinations. These results imply that similar mechanisms may govern cross-modal interactions across sensory modalities.

## 1 Introduction

In the course of our everyday activities, we are constantly exposed to an array of stimuli that activate more than one sensory modality. From this array, we may try to focus on some stimuli while trying to ignore or inhibit others. This ability is referred to as selective attention. This article deals with cross-modal selective attention, where the perceiver tries to focus on information in one sensory modality (the attended modality) and ignore information in another (the unattended modality). An extreme example of this ability is captured by the performance of Kerri Strug who helped the USA team win a gold medal in women’s gymnastics during the 1996 Olympics. Despite severe tendon and ligament damage following a first try at a vault, Strug managed to land a nearly perfect second vault by focusing on the sight of the apparatus while ignoring severe pain in her leg.

In general, research on selective attention has concentrated on the visual and auditory modalities. Unlike these previous studies, and consistent with the example above, we examined selective attention in vision and touch. We chose vision and touch because we wanted to determine whether processes hypothesized to explain cross-modal selective attention in vision and hearing could apply also to vision and touch. In this way we address a major issue in the study of cross-modal selective attention: do similar mechanisms underlie cross-modal selective attention across different sensory systems? We hypothesize that it may be impossible to attend to input from one modality while ignoring wholly input from another, and that common mechanisms govern failures of cross-modal selective attention. This line of reasoning follows from the general view that the senses are better conceptualized as interrelated modalities than as separate, independent channels (Marks 1978; see also Cytowic 1995 for a related claim).

### 1.1 *The cross-modal selective attention paradigm*

In one type of cross-modal selective attention task, participants are presented compound stimuli and are asked to classify values on an attended stimulus dimension under two conditions. In a *baseline* condition, values on the attended dimension vary from trial to trial while values on an irrelevant, unattended dimension remain constant.

In an *orthogonal* condition, values on both the attended and the irrelevant, unattended dimension vary independently. For example, in studies of selective attention to auditory pitch and visual lightness, participants are asked to classify a tone as high or low in pitch in the presence of a black or white patch of color. In different conditions, the lightness either remains constant (baseline) or varies randomly from trial to trial (orthogonal). In a complementary task, lightness serves as the attended dimension and pitch serves as the unattended one (eg Marks 1987; Martino and Marks 1999b; Melara 1989). Under these conditions, participants are unable to attend to one dimension without suffering ‘intrusion’ (ie a decrease in processing efficiency) from the other. This main result has been shown for several combinations of visual and auditory stimulus dimensions (see Martino and Marks 1999b, for relevant references).

### 1.2 *Failures of selective attention*

Using the paradigm just described, researchers have identified two types of intrusion that tend to co-occur. The first, Garner interference, is characterized by poorer overall performance under trial-by-trial orthogonal variation of irrelevant stimulus dimensions relative to baseline (eg Garner 1974; Garner and Felfoldy 1970). Garner interference is thought to reflect cross-talk during perceptual processing (eg Garner 1974; Melara and O’Brien 1987; Pomerantz 1983, 1986).

Of greater interest for our study is the second type of intrusion—congruence effects (eg Melara and O’Brien 1987). Congruence effects entail superior performance when values on attended and unattended dimensions ‘match’ rather than ‘mismatch’, and these effects generally occur when both dimensions vary orthogonally and not at baseline. In our studies, matches and mismatches reflect a synesthetic correspondence, or an intrinsic grouping, between stimuli on attended and unattended dimensions in different sensory modalities (see, for example, Martino and Marks 1999a). These groupings are determined by cross-modal rating or matching tasks (see also Melara and O’Brien 1987).

In the case of auditory pitch and visual lightness, participants group high pitches with white colored shapes and low pitches with black colored shapes [see Marks (1978) for additional examples]. Studies of selective attention to pitch and lightness show that when both dimensions vary orthogonally, congruence effects result. That is, when attending to pitch, participants are faster at classifying high pitches when these are accompanied by white (vs black) visual stimuli. Furthermore, participants are faster at identifying low pitches when these are accompanied by black (vs white) visual stimuli (Marks 1987; Martino and Marks 1999b; Melara 1989). Analogous results appear when participants try to attend to lightness while pitch varies orthogonally (eg Melara 1989). Such findings suggest that intrinsic groupings based on synesthetic correspondence are important in determining whether two dimensions interact. Consistent with this idea, dimensions that fail to show correspondence in rating or matching tasks do not produce congruence effects under orthogonal variation of stimulus dimensions (see Marks 1987).

In sum, studies of visual–auditory selective attention have identified two factors—orthogonal variation and synesthetic correspondence—important for predicting when selective attention will fail. These failures are useful indicators of cross-modal interaction during ‘on-line’ decision making. What is not yet clear is whether selective attention can fail in vision and touch as it does in vision and hearing, and by implication, whether the visual and tactile modalities interact.

### 1.3 *Two hypotheses regarding the locus of cross-modal interaction*

Two hypotheses have been advanced to explain why selective attention breaks down and where cross-modal interaction occurs. According to a physiognomic hypothesis (eg Werner 1957), perceptual stimuli produce modality-independent physiological responses that mediate cross-modal interaction. Since its inception, this hypothesis has been invoked

to explain a variety of phenomena including the tendency to associate different hues and brightness with mood states (D'Andrade and Egan 1994; Frank and Gilovich 1988).

Central to the physiognomic hypothesis is the proposal that there are interactions between sensory stimuli that occur at a 'low' level, separate from and prior to linguistic or semantic processes. That is, the hypothesis would attribute cross-modal interaction to early perceptual processing, prior to the engagement of linguistic or semantic mechanisms. This interpretation is consistent with reports that prelinguistic infants show pitch–position and brightness–loudness correspondences (see Braaten 1993; Lewkowicz and Turkewitz 1980; Wagner et al 1981).

The physiognomic hypothesis has trouble, however, in explaining the full pattern of cross-modal interaction described in the literature. In particular, according to the hypothesis, relationships between stimulus values from different modalities are defined absolutely. As a result, physiological responses should be evoked automatically not only under orthogonal variation of stimuli but also at baseline. This prediction is generally not borne out by the data. Congruence effects appear only with orthogonal variation of the stimulus attributes.

An alternative to the physiognomic hypothesis is the semantic coding hypothesis (see Martino and Marks 1999b). The semantic coding hypothesis makes three major claims. First, while failures of cross-modal selective attention may be due to sensory-level mechanisms in children, in adults these failures may reflect post-sensory (semantic) mechanisms created as a result of sensory and linguistic experience. Second, congruence effects reflect the coding or recoding of perceptual information into a common, abstract semantic representation that captures the synesthetic relation between dimensions. Third, the fact that congruence effects emerge only when dimensions vary orthogonally is attributed to the importance of trial-by-trial variation in stimulus values—only orthogonal variation provides a context in which to define *relative* stimulus values and thereby to highlight synesthetic relationships between dimensions.

In line with the semantic coding hypothesis, congruence effects can occur between dimensions that share verbal labels and synesthetic meanings (eg high/low pitch and position), and between dimensions that do not share labels, but do share synesthetic features (eg high/low pitch and white/black lightness)—see Ben-Artzi and Marks 1999; Martino and Marks 1999b; Melara 1989; Melara and O'Brien 1987. Additionally, recent investigations show that linguistic stimuli are sufficient to drive or modify congruence effects (eg high/low pitch, bigrams HI/LO) (Ben-Artzi and Marks 1999; Martino and Marks 1999b; Walker and Smith 1984). This last finding highlights the difference between the physiognomic and semantic coding hypotheses. The physiognomic hypothesis does not predict that linguistic stimuli will produce congruence effects because these effects should be mediated by low-level mechanisms. In contrast, the semantic coding hypothesis explains this finding as a result of encoding the sensory and linguistic stimuli alike into a post-perceptual representation that captures synesthetic correspondence between the dimensions.

## 2 Present experiments

For the present study, we developed a selective attention task, similar to the one described earlier, to study interactions between vision and touch. In this way, we hoped to discern whether common mechanisms underlie cross-modal interaction and selective attention across sensory modalities (in vision–audition and vision–touch). In addition to this broad aim, we developed three hypotheses. First, given the semantic coding hypothesis, we predicted that, under orthogonal variation of stimulus dimensions, performance should be superior when visual and tactile stimuli match, than when they mismatch, a congruence effect. To test this prediction, we chose the dimensions of visual lightness (black, white) and vibrotactile frequency (low, high). We confirmed the synesthetic

correspondence between these dimensions in a pilot study described elsewhere (Martino and Marks 1999a). Based on this pilot study, matches were defined as black + low frequency and white + high frequency and mismatches were defined as black + high frequency and white + low frequency. Second, given results from analogous studies of pitch and lightness interaction, we predicted that vision and touch should show Garner interference; that is, performance should be poorer under orthogonal variation of stimuli than at baseline. Finally, we expected congruence effects to be bidirectional, occurring when participants attend to vision and when they attend to touch.

## 2.1 Method

**2.1.1 Participants.** Thirteen men and fifteen women performed vibration classification, and twenty men and sixteen women performed lightness classification. All had normal or corrected-to-normal vision. Participants were compensated \$8.00 per hour for volunteering.

**2.1.2 Stimuli and apparatus.** Two sinusoids (500 ms each) matched for perceived intensity were created ('high' vibration 200 Hz; 'low' vibration 50 Hz). These stimuli were generated by computer and presented via a shaker (Brüel & Kjaer) to a stylus whose 1 cm diameter tip was in contact with the skin. The stylus extended up through a 1.2 cm diameter hole, leaving a 1 mm gap between stylus and the rigid surround. The shaker was mounted on a triple-beam balance to ensure that the tip of the stylus would exert a constant pressure of 30 g on the thenar eminence of the participant's hand. To prevent participants from hearing the vibratory stimuli, flat-spectrum noise, filtered to exclude frequencies above 2400 Hz, was presented via circumaural headphones.

Black and white squares (5.1 cm × 5.1 cm) were created with Superpaint (Aldus) (black squares: hue angle 180°, saturation 0%, and lightness 0%; white squares: hue angle 180°, saturation 0%, and lightness 100%). The squares were presented centrally against a gray background (hue angle 288°, saturation 0%, and lightness 37.5%) on a 35 cm color monitor. The luminances of the white and black stimuli were 40 and 1.2 cd m<sup>-2</sup>, respectively, and the luminance of the gray background was 8.1 cd m<sup>-2</sup>.

A trial consisted of a vibratory stimulus and a visual stimulus presented simultaneously. Trials were response-terminated. Superlab (Cedrus) was used to coordinate the presentation and randomization of trials and to time the responses.

**Vibration classification.** In the baseline condition, the vibration frequency varied randomly from trial to trial while the lightness of the visual stimulus remained constant. There were two baseline tasks (lightness: white or black) of 48 trials each. In the orthogonal condition, both lightness and vibration varied randomly and independently (96 trials). There were equal numbers of trials of each stimulus combination. The order of baseline and orthogonal tasks was counterbalanced across participants. Each task was preceded by 20 practice trials.

**Lightness classification.** Lightness classification paralleled vibration classification. The only change occurred in the baseline condition, in which lightness varied from trial to trial while the vibratory frequency remained constant.

**2.1.3 Procedure.** Participants were tested individually for ~½h in a sound-attenuated chamber. On each trial, participants saw a visual stimulus and felt a vibratory stimulus simultaneously. Participants who classified the vibratory stimuli did so by pressing, as quickly as possible, one key if the vibration was low and another key if the vibration was high. Participants who classified the visual stimuli pressed, as quickly as possible, one key if the square was black and another if it was white. All participants logged their responses with their right hand using their index or second finger (counterbalanced across participants) and felt the vibrations with their left hand. When not responding, participants rested their fingers on the response keys.

Participants were encouraged to look at the computer monitor even though they were not asked to respond to the visual stimuli. To help participants direct their gaze appropriately, they rested their head against a combination head-and-chin rest  $\sim 46$  cm from the computer screen.

## 2.2 Results

**2.2.1 Treatment of data.** To prepare the data for the analyses to be described below, we excluded incorrect responses and trimmed outliers. An outlier was defined as any response time greater than two SDs from the mean of the cell for that person (the SD was calculated with outliers included). Outliers were trimmed by replacing the extreme value with the value two SDs from the mean. The average number of trimmed reaction times for vibration classification was 7.4 (3.8%) with a range of 3 to 10. Similar values were obtained for lightness classification, 6.6 (3.4%) with a range of 3 to 13.

To assess the possibility that extreme scores may exert a disproportionate influence on an analysis of variance, we first analyzed arithmetic means and then analyzed logarithmically transformed (or geometric) means, as described by Kirk (1982). The pattern of results did not change as a result of this transformation and thus we report analyses based on transformed latencies only.

A breakdown of the average number of errors across tasks and conditions is shown in table 1. Accuracy was nearly perfect, so we did not analyze errors except to investigate possible speed–accuracy tradeoffs. Pearson’s  $r$  correlations between average logarithm of response time (log RT) and average number of errors showed no reliable speed–accuracy tradeoffs for vibration or lightness classification (vibration classification, baseline:  $r_{26} = -0.3$ ; orthogonal:  $r_{26} = -0.2$ ; lightness classification, baseline:  $r_{34} = -0.2$ ; orthogonal:  $r_{34} = -0.2$ , all ns).

**2.2.2 Vibration classification.** A condition (baseline, orthogonal)  $\times$  congruence (match, mismatch) analysis of variance (ANOVA) revealed no main effect of condition, which would have been consistent with a significant effect of Garner interference. Thus, responses were no faster overall in the baseline condition than in the orthogonal condition ( $F < 1.0$ ).

**Table 1.** Mean reaction times (RTs) and errors in classifying vibrations or lightness as a function of task condition (baseline, orthogonal) and congruence (match, mismatch). Separate groups of participants performed vibration classification and lightness classification.  $\Delta_{b-o} = M_{\text{baseline}} - M_{\text{orthogonal}}$ ;  $\Delta_{m-n} = M_{\text{match}} - M_{\text{mismatch}}$ .

	Overall means		Baseline means		Orthogonal means				
	baseline	orthogonal	$\Delta_{b-o}$	match	mismatch	$\Delta_{m-n}$	match	mismatch	$\Delta_{m-n}$
Vibration classification									
RT/ms	582.0	574.0	+8.0	582.0	582.0	0.0	566.0	582.0	-16.0
errors	0.5	0.7	-0.2	0.5	0.5	0.0	0.7	0.7	0.0
Lightness classification									
RT/ms	431.5	436.5	-5.0	435.0	428.0	+7.0	433.0	440.0	-7.0
errors	0.5	0.5	0.0	0.5	0.6	-0.1	0.5	0.5	0.0

Note: A negative  $\Delta_{b-o}$  suggests that the baseline tasks were performed more efficiently than the orthogonal task, a pattern indicative of Garner interference. We do not observe significant amounts of Garner interference for vibration or lightness classification. A negative  $\Delta_{m-n}$  reflects faster RT to matches versus mismatches. The semantic coding hypothesis predicts that under orthogonal variation, but not under baseline, RT to matches will be faster than for mismatches (a congruence effect). Congruence effects were statistically significant for both vibration and lightness classification.

There was, however, an interaction of the two factors,  $F_{1,27} = 5.9$ ,  $MSE = 0.001$ ,  $p < 0.05$ . Planned contrasts confirmed a pattern typical of a congruence effect. In the orthogonal condition, matching stimuli were identified 16 ms more quickly than were mismatching stimuli, a significant difference,  $F_{1,27} = 11.6$ ,  $MSE = 0.002$ ,  $p < 0.05$ . There was no significant difference between mean latencies for matches and mismatches in the baseline condition,  $F_{1,27} < 1.0$ . The magnitude of this congruence effect is generally consistent with those reported for vision–audition interactions (eg Melara 1989).

We performed an additional comparison to explore why the congruence effect occurred—was it produced by facilitation from matching pairs, interference from mismatching pairs, or both? This comparison suggested response facilitation: Relative to baseline, participants responded quickly to matching stimuli (orthogonal condition),  $F_{1,27} = 14.8$ ,  $MSE = 0.003$ ,  $p < 0.05$ ; but neither more nor less quickly to mismatching stimuli,  $F_{1,27} < 1.0$ . The means for these comparisons appear in table 1.

**2.2.3 Lightness classification.** As with the findings for vibration classification, an analogous ANOVA on latencies to classify lightness showed (a) no main effect of condition, implying a lack of significant Garner interference,  $F < 1.0$ ; but (b) an interaction of condition with congruence,  $F_{1,35} = 8.4$ ,  $MSE = 0.002$ ,  $p < 0.05$ . Planned contrasts showed that in the orthogonal condition, participants responded significantly more quickly to matching stimuli than to mismatching ones,  $F_{1,35} = 4.2$ ,  $MSE = 0.001$ ,  $p < 0.05$ . The same analysis revealed an effect in the opposite direction in the baseline condition,  $F_{1,35} = 4.2$ ,  $MSE = 0.001$ ,  $p < 0.05$ .

Other planned contrasts showed that, unlike the results obtained when participants classified vibration, when participants classified lightness the congruence effect implied interference from mismatches and not facilitation from matches (baseline vs mismatch:  $F_{1,35} = 7.6$ ,  $MSE = 0.002$ ,  $p < 0.05$ ; baseline vs match:  $F_{1,35} < 1.0$ ).

**2.2.4 Vibration classification versus lightness classification.** To test which classification task was performed more efficiently overall, we conducted a task (vibration or lightness classification)  $\times$  condition  $\times$  congruence ANOVA. This analysis showed that participants classified lightness 144 ms more quickly on average than they classified vibration (578 ms versus 434 ms),  $F_{1,62} = 14.3$ ,  $MSE = 0.97$ ,  $p < 0.001$ . Consistent with the analyses above, there was an interaction of condition with congruence, such that, in the orthogonal condition, matching stimuli were identified 12 ms more quickly on average than were mismatching stimuli (499 ms versus 511 ms),  $F_{1,62} = 14.7$ ,  $MSE = 0.003$ ,  $p < 0.001$ . The reverse pattern was observed in the baseline condition, with slightly but not reliably longer latencies for matches (508 ms) than mismatches (505 ms),  $F_{1,62} = 2.0$ , ns ( $F_{1,62} = 13.8$ ,  $MSE = 0.003$ ,  $p < 0.001$  for the two-way interaction term).

### 3 Discussion

Consistent with the hypothesis that there is a fundamental “unity of the senses”, one important result of this study is the general finding that participants could not attend wholly to stimulation of one modality without intrusion from stimulation of another modality. Both when participants attended to touch and when they attended to vision, performance was affected by activity in the unattended channel. This result is consistent with a number of findings reviewed by Pashler (1998), showing that unattended stimuli can be processed at least partially, thereby influencing the processing of attended stimuli.

Contrary to our expectations, the pattern of results obtained here with vision and touch did not mirror precisely the pattern reported previously with vision and hearing. Consistent with our predictions, under orthogonal variation of stimulus dimensions participants responded more quickly to matches than to mismatches. This pattern emerged when participants classified vibrations and when they classified

lightness (a bidirectional effect), thereby providing the first demonstration of congruence effects for visual and tactile dimensions. The finding is consistent with analogous studies of interactions between visual lightness and auditory frequency. However, unlike findings in vision and hearing, the congruence effects observed here displayed clear evidence of facilitation as well as interference and were not accompanied by a significant amount of Garner interference. We discuss these major results, and their implications, below.

### 3.1 Congruence effects

The finding that trial-by-trial variation of cross-modally related stimulus values affects one's ability to attend selectively cannot be explained by the physiognomic hypothesis. Such a hypothesis would predict that responses should be evoked automatically, whenever compound stimuli are presented. Thus, the same pattern of results should appear in both the baseline and orthogonal conditions—with participants responding more rapidly to matching than mismatching stimuli. This did not occur. In fact, when classifying lightness at baseline, participants responded slightly more rapidly to mismatches than to matches.

The data appear more compatible with the semantic coding hypothesis. This view maintains that, over years of linguistic experience, meaningful relationships among perceptual, nonlinguistic stimulus dimensions are established. Under trial-by-trial variation of stimulus dimensions, these relations become salient and affect the ability to classify stimuli on one dimension without significant intrusion from the other (see also Martino and Marks, in press).

According to the semantic coding hypothesis, the locus of these cross-modal interactions is post-perceptual, occurring after stimulus information is coded or recoded into an abstract semantic representation that captures the synesthetic correspondence common to dimensions of both modalities. In the case of vision (eg lightness) and touch (eg vibration), this representation is perhaps built from experience with textures gleaned through vision, roughness gleaned through touch, and the labels we use to describe these visual and tactile experiences.

It is possible that verbal labels per se, rather than abstract representations, underlie cross-modal dimensional interaction? Some have argued that, when dimensions share verbal labels, cross-modal dimensional interaction could occur at the level of verbal coding. For example, this argument is given to explain cross-modal dimensional interaction between auditory pitch and vertical position because both dimensions share the labels 'high' and 'low' (eg Melara and O'Brien 1987). Unlike the dimensions pitch and position, vibration frequency and visual lightness do not share commonly used verbal labels. This being the case, we believe our findings cannot be explained satisfactorily by a verbal coding explanation.

Note that the size of the congruence effect was larger when participants classified vibration compared to lightness—an outcome of interest given our finding that lightness was processed more quickly overall (ie was more 'discriminable') than was vibration. This pattern is generically consistent with an explanation in terms of speed of processing. According to this explanation, when two dimensions are processed at different rates, the dimension processed more quickly is likely to affect the dimension processed more slowly, but not vice versa (eg Morton and Chambers 1973; Posner and Snyder 1975). This explanation is often used to account for the Stroop effect (Stroop 1935; see also MacLeod 1991).

If our two dimensions were equally discriminable (eg made equal in RT at baseline) would the results change? This question stems from the finding that some instances of dimensional interaction are modulated by the speed at which one can make discriminations among dimensional values. For example, Arieh and Algom (under review)

report that Stroop interactions and Garner interference between words and pictures diminish when decision times for words and pictures are matched at baseline (see also Melara and Mounts 1994). At this point, it is not clear whether cross-modal interactions are affected by differences in baseline RT as are within-modality interactions (eg Stroop effects) because a systematic study of this issue has not yet been performed. Congruence effects have been reported for visual and auditory dimensions roughly 'matched' (< 50 ms difference in RT) and 'unmatched' (> 50 ms difference in RT) at baseline (see Martino and Marks 1999b for a list of relevant references). Further research is necessary to test whether changes in baseline discriminability would change the pattern of congruence effects and/or Garner interference observed for visual–auditory and visual–tactile dimensions.

### 3.2 *Interference and facilitation in congruence effects*

Although the semantic coding hypothesis provides a generic account of when and where congruence effects occur, it leaves open whether the effects represent facilitation from matched pairings, interference from mismatched pairings, or both. Our statistical results appear equivocal. When participants classified vibration, performance suggested facilitation; but when they classified lightness, performance suggested interference. We believe that an information-accrual model described by Ben-Artzi and Marks (1995) and Martino and Marks (1999b) can help clarify this matter.

Following a suggestion of Ben-Artzi and Marks (1995), Martino and Marks (1999b) describe an 'addendum' to the semantic coding hypothesis to address the question of how failures of selective attention occur, particularly when the two dimensions are synesthetically related.

This model makes three claims. First, sensory information presented to two modalities accrues over time toward a classification threshold. Initially, the information is processed in parallel via separate sensory channels. Second, when the two dimensions are synesthetically related (or, in some cases, literally related), this perceptual information becomes coded or recoded into an abstract semantic representation at a post-perceptual locus per the semantic coding hypothesis. Third, two independent factors affect the efficiency with which information accumulates: the manner in which the irrelevant dimension is presented (ie constant versus orthogonal variation), and the relationship between the relevant and irrelevant dimensions (ie synesthetic congruence). These two factors are thought to aid or hinder the accrual of information, leading to 'fast' or 'slow' decision times.

In the case of orthogonal variation, the trial-by-trial change along the irrelevant dimension hinders information accrual toward a decision threshold, leading to Garner interference. Just how this occurs is unclear—it may be that orthogonal stimulus variation affects the location of a response criterion (although this could entail a corresponding decrease in errors), diverts some attentional resources toward the irrelevant dimension, or both, depending on task demands.

Now suppose that values on the relevant and irrelevant dimensions match synesthetically. In principle, information accrued on the irrelevant dimension could aid classification by functionally 'adding to' information accrued on the critical dimension, leading to response facilitation. The opposite could occur when attributes mismatch synesthetically, leading to response interference.

This effect of synesthetic correspondence is predicted to occur under orthogonal variation only. This is because trial-by-trial variation on the irrelevant dimension establishes a context with which to define the stimulus values as matches and mismatches.

Given the possibility that Garner interference and congruence effects arise from separate sources, one may or may not see the effects co-occur. If, as we observed here, significant congruence effects occurred without significant Garner interference, the model



could accommodate this observation in two ways. First, it may be that interference due to orthogonal variation relative to baseline is completely absent. This means that the irrelevant dimension did not hinder information accrual as expected. This could occur, for example, if the two dimensions were separable, as described by Garner (1974). When dimensions are separable, by definition attention can be directed to one dimension wholly without intrusion from another. Such a result has been documented for the visual dimensions size and brightness (Garner 1977). Second, it may be that Garner interference is indeed present, but that facilitation from congruent trials and interference from incongruent trials summed in such a way as to cancel any overall effect of Garner interference. In the case of vibration, facilitation may have been greater than interference. In the case of lightness, interference may have been greater than facilitation. In both cases, however, these effects may have cancelled any overall interference due to orthogonal variation. At present, this model is not detailed enough to allow us to choose between these two possible explanations of our results.

The presence of congruence effects without significant amounts of Garner interference differs from the pattern typically observed in comparable studies of vision and hearing, in which both effects are evident (eg Martino and Marks 1999b; Melara 1989). Thus, selective attention with synesthetically corresponding visual and tactile dimensions reveals a new profile of cross-modal interaction. Like our study of vision and touch, comparable studies of vision and hearing are consistent with two versions of the information-accrual model. In one version, interference from orthogonal variation itself and interference from stimulus mismatches combine to produce Garner interference and congruence effects, respectively. In a second version, interference from orthogonal variation combines not only with interference from incongruent pairings but also with facilitation from stimulus matches.

At this time, we lack a formal theory to describe the mechanisms of cross-modal interaction that lead to failures of selective attention across sensory modalities. Nevertheless, in service of this goal, we demonstrate that failures of selective attention for vision and touch, like those previously documented for vision and hearing, are consistent with the semantic coding hypothesis and are usually interpreted within an information-accrual model. One implication of these findings is clear: The processes underlying cross-modal interactions between vision and touch are likely to be similar, although not identical, to the processes underlying interactions between vision and hearing.

**Acknowledgements.** This work was supported by NIH Postdoctoral Training Fellowship T32 DC00025-13 to the first author and NIH grant R01 DC02752-02 to the second author. We thank Sarah Marshall and Leah Berman for their assistance.

## References

- Arieh Y, Algom D, (under review) "Selective attention with picture–world stimuli: neither superiority for word or picture nor Stroop and Garner effects are inevitable" *Journal of Experimental Psychology: Learning, Memory and Cognition*
- Ben-Artzi E, Marks L E, 1995 "Congruence effects in classifying auditory stimuli: a review and a model", in *Proceedings of the Eleventh Annual Meeting of the International Society for Psychophysics* Ed. C A Possamai (Cassis, France: International Society for Psychophysics) pp 145–150
- Ben-Artzi E, Marks L E, 1999 "Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification" *Journal of Experimental Psychology: Human Perception and Performance* **25** 1–17
- Braaten R, 1993 "Synesthetic correspondence between visual location and auditory pitch in infants", paper presented at the Annual Meeting of the Psychonomic Society, Washington, DC, November
- Cytowic R E, 1995 *The Neurological Side of Neuropsychology* (Cambridge, MA: MIT Press)
- D'Andrade R, Egan M, 1974 "The colors of emotions" *American Ethnologist* **1** 49–63
- Frank M G, Gilovich T, 1988 "The dark side of self- and social perception: black uniforms and aggression in professional sports" *Journal of Personality and Social Psychology* **54** 74–85
- Garner W R, 1974 *The Processing of Information and Structure* (Hillsdale, NJ: Lawrence Erlbaum Associates)

- Garner W R, 1977 "The effect of absolute size on the separability of the dimensions of size and brightness" *Bulletin of the Psychonomics Society* **9** 380–382
- Garner W R, Felfoldy G L, 1970 "Integrality of stimulus dimensions in various types of information processing" *Cognitive Psychology* **1** 225–241
- Kirk R E, 1982 *Experimental Design* 2nd edition (Monterey, CA: Brooks Cole)
- Lewkowicz D J, Turkewitz G, 1980 "Cross-modal equivalence in early infancy: Auditory–visual intensity matching" *Developmental Psychology* **16** 597–607
- MacLeod C, 1991 "Half a century of research on the Stroop effect: an integrative review" *Psychological Bulletin* **109** 163–203
- Marks L E, 1978 *The Unity of the Senses: Interrelations among the Modalities* (New York: Academic Press)
- Marks L E, 1987 "On cross-modal similarity: auditory–visual interactions in speeded discrimination" *Journal of Experimental Psychology: Human Perception and Performance* **13** 384–394
- Martino G, Marks L E, 1999a "Cross-modal interaction in vision and touch", poster session presented at the 40th Annual Meeting of the Psychonomics Society, Los Angeles, CA, November
- Martino G, Marks L E, 1999b "Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis" *Perception* **28** 903–923
- Martino G, Marks L E, in press, "Synesthesia: Strong and weak" *Current Directions in Psychological Science*
- Melara R D, 1989 "Dimensional interaction between color and pitch" *Journal of Experimental Psychology: Human Perception and Performance* **115** 212–231
- Melara R D, Mounts J W, 1994 "Contextual influences on interactive processing: effects of discriminability, quantity, and uncertainty" *Perception & Psychophysics* **56** 73–90
- Melara R D, O'Brien T P, 1987 "Interaction between synesthetically corresponding dimensions" *Journal of Experimental Psychology: General* **116** 323–336
- Morton J, Chambers S M, 1973 "Selective attention to words and colours" *Quarterly Journal of Experimental Psychology* **25** 387–397
- Pashler H E, 1998 *The Psychology of Attention* (Cambridge, MA: MIT Press)
- Pomerantz J R, 1983 "Global and local precedence: Selective attention in form and motion perception" *Journal of Experimental Psychology: General* **112** 515–540
- Pomerantz J R, 1986 "Visual form perception: an overview", in *Pattern Recognition by Humans and Machines: Visual Perception* Eds E C Schwab, H C Nusbaum (New York: Academic Press) pp 1–30
- Posner M I, Snyder C R R, 1975 "Attention and cognitive control", in *Information Processing and Cognition: The Loyola Symposium* Ed. R L Solso (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 55–85
- Stroop J R, 1935 "Studies of interference in serial verbal reactions" *Journal of Experimental Psychology* **18** 643–662
- Wagner S, Winner E, Cicchetti D, Gardner H, 1981 "'Metaphorical' mapping in human infants" *Child Development* **52** 728–731
- Walker P, Smith S, 1984 "Stroop interference based on the synesthetic qualities of auditory pitch" *Perception* **13** 75–81
- Werner H, 1957 *Comparative Psychology of Mental Development* revised edition (New York: International Universities Press)