Multisensory synesthetic interactions in the speeded classification of visual size

ALBERTO GALLACE
University of Oxford, Oxford, England
and Università degli Studi di Milano Bicocca, Milan, Italy

and

CHARLES SPENCE
University of Oxford, Oxford, England

In the present study, we attempted to demonstrate a synesthetic relationship between auditory frequency and visual size. In Experiment 1, participants performed a speeded visual size discrimination task in which they had to judge whether a variable-sized disk was bigger or smaller than a standard reference disk. A task-irrelevant sound that was either synesthetically congruent with the relative size of the disk (e.g., a low-frequency sound presented with a bigger disk) or synesthetically incongruent with it (e.g., a low-frequency sound presented with a smaller disk) was sometimes presented together with the variable disk. Reaction times were shorter in the synesthetically congruent condition than in the incongruent condition. Verbal labeling and semantic mediation interpretations of this interaction were explored in Experiment 2, in which high- and low-frequency sounds were presented in separate blocks of trials, and in Experiment 3, in which the tones were replaced by the spoken words "high" and "low." Response priming/bias explanations were ruled out in Experiment 4, in which a synesthetic congruency effect was still reported even when participants made same-versus-different discrimination responses regarding the relative sizes of the two disks. Taken together, these results provide the first empirical demonstration that the relative frequency of an irrelevant sound can influence the speed with which participants judge the size of visual stimuli when the sound varies on a trial-by-trial basis along a synesthetically compatible dimension. The possible cognitive bases for this synesthetic association are also discussed.

The speeded classification paradigm has long been used by psychologists to study a variety of issues in selective attention research (e.g., Garner, 1974; see Marks, 2004, for a recent review). In a typical study, participants have to discriminate one characteristic of a stimulus (e.g., its size) as rapidly as possible while trying to ignore any "irrelevant" characteristics of the stimulus (e.g., its brightness) that may also vary on a trial-by-trial basis (see, e.g., Garner, 1977). Many studies over the last 30 years have shown that a participant's response to the relevant characteristic of a stimulus can be influenced by variations in the irrelevant stimulus dimension, giving rise to what has become known as *Garner interference* (e.g., Garner, 1974; Pomerantz, Pristach, & Carson, 1989) or *congruence effects* (e.g., Clark & Brownell, 1975; Patching & Quinlan, 2002).[1]

These effects have also been demonstrated to occur cross-modally when the relevant and irrelevant stimulus characteristics are presented in different sensory modalities (see, e.g., Bernstein & Edelstein, 1971; Taylor & Campbell, 1976; see Marks, 2004, for a review). For example, Bernstein and Edelstein asked participants to discriminate the side of a display on which a visual stimulus

was presented (e.g., left or right) while trying to ignore a simultaneously presented monaural tone. The participants responded more rapidly on trials on which the two stimuli were presented on the same side than on those trials on which they were presented on different sides (but see Spence & McDonald, 2004).

The speeded classification paradigm has also been used across synesthetically defined cross-modal dimensions.[2] For instance, Melara and O'Brien (1987) reported a series of experiments in which they explored whether synesthetic correspondences in nonsynesthetic individuals were based on perceptual (rather than allegorical or inferential) interactions. Participants were presented simultaneously with a sequence of visual stimuli whose elevation (e.g., higher vs. lower) varied and a low- or high-frequency tone (174.6 vs. 1046.5 Hz, respectively; see also Pratt, 1930; Stumpf, 1883). Response latencies to classify the elevation of the visual stimuli were longer when the irrelevant tones were synesthetically incongruent with the target (e.g., a higher light presented with a lower frequency sound) than when they were synesthetically congruent (e.g., a higher visual stimulus presented with a higher frequency sound; see also Bernstein & Edelstein, 1971). Melara and O'Brien argued that the matching of cross-modal attributes may have made it easier for their participants to distinguish between two stimuli formed from pairs of synesthetically congruent attributes than between stimuli formed from pairs of incon-

gruent attributes. That is, stimuli formed from synesthetically congruent attributes are more likely to group together, and thus to become more easily classifiable, than stimuli formed from synesthetically incongruent attributes (e.g., it is easier to distinguish the stimulus pairs "( )" and "[ ]" than the pairs "[ )" and ") ["). Melara and O'Brien proposed a number of possible explanations for the synesthetic congruency effects highlighted in their study (e.g., response competition, mediation by verbal codes, perceptual integrality and/or configurality) but concluded that no single explanation was sufficient to account for all of their data.

Congruency effects involving a synesthetically related dimension have also been demonstrated between vision and touch. Participants in Martino and Marks's (2000) study had to judge the perceived lightness (black vs. white) of a sequence of visual stimuli while ignoring a low- or high-frequency vibrotactile stimulus presented to their fingertips or vice versa (i.e., judge the frequency of vibrotactile target stimuli while ignoring visual distractor stimuli). Their participants responded most rapidly to high-frequency vibrations accompanied by white stimuli and low-frequency vibrations accompanied by black stimuli. Martino and Marks's (2000) results therefore highlight the existence of a cross-modal congruency effect that is dependent on the existence of a synesthetic relationship between vision and touch.

Synesthetic correspondences between audition and touch in nonsynesthetes have also been studied. Walker and Smith (1985) reported that reaction times (RTs) to classify words as either "little" or "big" were longer when the pitch of a simultaneously presented sound and/or the size of a simultaneously presented knob handle was incongruent with the multisensory features represented by the test word (e.g., when the word *little* was presented together with a low-pitched sound and/or a large handle;[3] see also Pratt, 1930). However, Walker and Smith argued that in their study haptically perceived size and auditory pitch were more likely to be interacting at a semantic, or response-selection, stage than at a perceptual level (see also Marks, 2004, on this point; cf. Rubinstein & Henik, 2002).

The aim of the present study was to explore whether or not the classification of the size of a visually presented disk would also be influenced by the presentation of an irrelevant auditory stimulus varying along a synesthetically related dimension—namely, frequency. If higher frequency tones are associated with small objects and lower frequency tones with large objects (as suggested by the results of Walker & Smith's, 1984, audiotactile study; see also Pratt, 1930, p. 280; Stumpf, 1883, p. 207), one would expect to find better performance for combinations of auditory and visual stimuli that are synesthetically congruent than for combinations that are incongruent. Exploring the possible relationship between the visual size of an object and the frequency of a simultaneously presented sound should help to answer some important questions concerning the nature of, and mechanisms underlying, synesthetic correspondences in nonsynesthetic individuals. In particular, it is important to try to determine whether synesthetic correspondences are limited to certain pairs of sensory modalities and/or to specific perceptual dimensions (e.g., brightness, elevation,

color), or instead reflect a more general equivalence of the senses underlying perception (see, e.g., Marks, 1978). A further important issue to be addressed by the present study concerns the role of verbal mediation in the elicitation of synesthetic correspondences in nonsynesthetic individuals.

EXPERIMENT 1

Method

Participants. Fifteen right-handed participants (7 males and 8 females; mean age, 26 years; range, 20–40 years) took part in Experiment 1 as paid volunteers. All of the participants reported normal auditory perception and normal or corrected-to-normal vision. The experiment took approximately 20 min to complete, and all of the participants received a £5 gift voucher in return for their participation. The experiments described in this study were noninvasive and had ethical approval from the Department of Experimental Psychology at the University of Oxford. The experiments were also performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus and Materials. The experiment was conducted in a dimly illuminated sound-attenuated booth. Each participant sat at a desk, 50 cm from a laptop computer with a 28.5×21.5 cm LCD screen. The visual stimuli were presented on the LCD screen, and the auditory stimuli were presented over two loudspeakers integrated into the PC and situated on both sides of the screen (center-to-center distance of 22 cm between the loudspeakers). Figure 1 illustrates the sequence of events of each trial. At the start of the trial, a red fixation point (0.3 cm in diameter) was presented in the middle of the screen for 300 msec and then disappeared, to be replaced by a blank white screen. Three hundred milliseconds later, a light gray (standard) disk (5.5 cm in diameter) was presented at the center of the screen, and after a further 300 msec this disk was replaced by a random dot mask that filled the entire screen, with dots ranging in brightness from light to dark gray. The visual mask was presented for 500 msec and was followed by a second light gray disk (of variable diameter) on the screen. The position of the second disk varied randomly between successive trials (ranging to ±0.3 cm vertically and horizontally from the center of the screen) to prevent the participant from using superimposition cues to determine any difference in the size of the two circles. The diameters of the variable-sized disks were ±5%, ±10%, ±20%, and ±40% of the diameter of the standard disk. The variable disk was presented on the screen for 80 msec and then obscured by a second visual mask, which stayed on the screen until a response had been made or until 3,000 msec had elapsed, at which time a new trial began. The intertrial interval was fixed at 300 msec. The variable disk was presented together with a low-frequency (300-Hz) tone, a high-frequency (4500-Hz) tone, or no tone. The tone was presented for 300 msec at an intensity of 75 dB. Each of the eight different variable-sized disks was presented 10 times in each of the three sound conditions (high frequency, low frequency, and no sound), giving rise to a total of 240 trials in each participant's experimental session.

The participants were asked to keep their feet pressed down on two foot pedals positioned 15 cm to either side of body midline, each under the toes of one foot. The participants were instructed to lift their toes off the left pedal if the second gray disk appeared to be larger than the first, and to lift their toes off the right pedal if it appeared to be smaller than the first. The participants were informed that a task-irrelevant sound would sometimes be presented at the same time as the variable disk, but they were asked to ignore it as much as possible. They were also informed that the standard disk would always be presented first. No feedback regarding the correctness of the participant's responses was provided at any point during the experiment. The presentation of the stimuli and the monitoring of the participants' responses were controlled by a 466-MHz processor laptop running MATLAB 6.1 (Release 12).
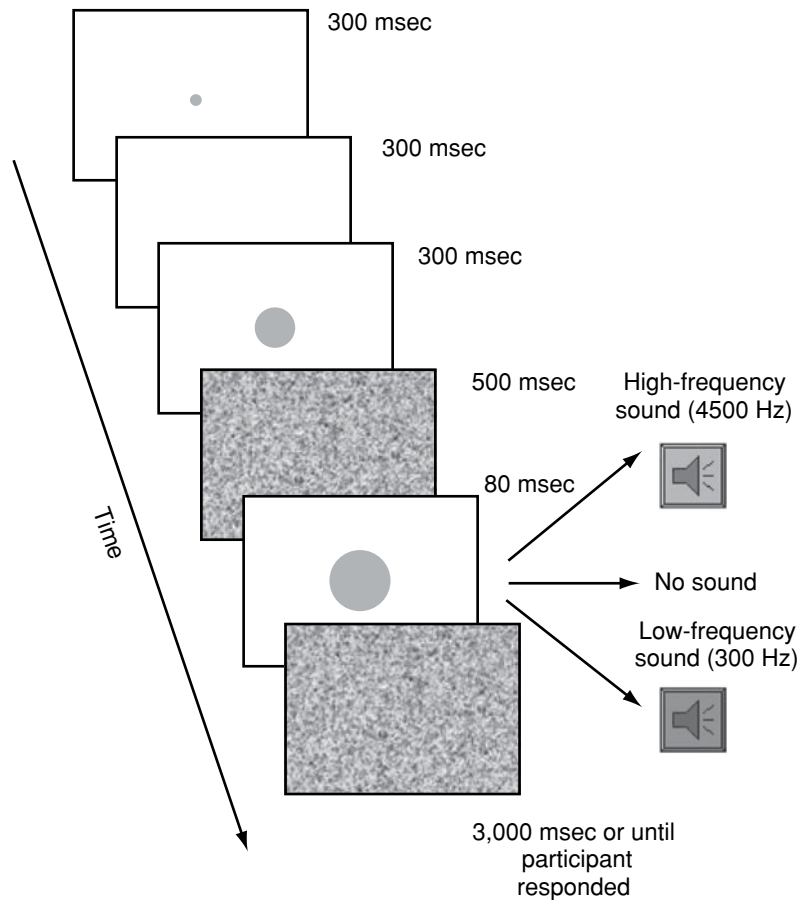
Figure 1. Schematic illustration of the sequence of events presented in each trial of Experiments 1, 3, and 4. The numerical values indicate the duration (in milliseconds) of each screen event.

Procedure. The participants were seated comfortably in the soundproof booth and asked to place their feet in the appropriate position, one over each foot pedal. The experimenter explained the instructions, and a series of practice trials was presented for 1 min to ensure that the participants had clearly understood the task. The experiment was composed of two blocks of 120 trials, with a short break between them. After the first block, the participants chose when to start the second block of trials by pressing a key on the keyboard. The participants were instructed to respond as rapidly as possible. Trials during which the participants failed to make a response (within 3,000 msec) were not analyzed. Each of the trials was classified as synesthetically congruent (i.e., a high-frequency sound presented with a variable disk that was smaller than the standard, or a low-frequency sound presented with a variable disk that was bigger than the standard), incongruent (i.e., a high-frequency sound presented with a variable disk that was bigger than the standard, or a low-frequency sound presented with a variable disk that was smaller than the standard), or neutral (i.e., no sound presented with the variable disk). Note that equal numbers of congruent, incongruent, and neutral trials were presented in all four experiments.

Results

The mean RT data from Experiment 1 (see Table 1) were submitted to an 8 × 3 repeated measures ANOVA with size of the variable stimulus ($\pm 5\%$ vs. $\pm 10\%$ vs. $\pm 20\%$ vs. $\pm 40\%$ of the diameter of the standard disk)

and sound (congruent vs. incongruent vs. no sound) as factors. The analysis revealed a significant main effect of size of the variable disk [$F(7,98) = 8.90, p < .001$], with response latencies decreasing as the difference in size between the standard and the variable disk increased. The analysis of the RT data also revealed a significant main effect of sound [$F(2,28) = 19.96, p < .00001$]. Post hoc least significant difference (LSD) tests revealed significant differences between all three sound conditions (all ps $< .05$), with the slowest responses occurring in the neutral (no-sound) condition and the fastest in the congruent condition (see Figure 2). There was no significant interaction between size and sound [$F(14,196) < 1$].

A similar ANOVA was performed on the error data. This revealed a significant main effect of size of the variable disk [$F(7,98) = 17.71, p < .0001$], with the error rate increasing as the difference between the sizes of the standard and variable disks decreased, as was expected. Although the participants made numerically the most errors in the incongruent sound condition and the fewest in the congruent sound condition (see Figure 2), the main effect of sound was not significant in the analysis of the error data [$F(2,28) = 1.58$]. The size × sound interaction

Table 1
Mean Reaction Times (RTs, in Milliseconds), Their Standard Errors, and Mean Error Rates for Participants
in Judging Each Stimulus Size in Experiments 1–4 As a Function of Task, Sound Condition,
and Size of the Variable Disk Relative to That of the Standard

			Sound Condition								
			Congruent			Incongruent			No Sound		
Task	Irrelevant Stimulus	Relative Size of Variable Disk	RT	SE	%E	RT	SE	%E	RT	SE	%E
			Experiment 1								
Size discrimination	sound	−40%	514	29	4.7	553	29	3.3	571	16	2.7
(bigger vs. smaller)		−20%	539	25	5.3	588	31	3.3	567	23	2.7
		−10%	588	42	6.7	608	32	10.7	631	28	8.0
		−5%	619	45	21.3	631	31	22.0	689	37	18.7
		+5%	647	33	18.7	651	30	32.0	666	41	31.3
		+10%	602	41	7.3	593	25	12.0	639	28	16.0
		+20%	526	25	4.0	554	32	6.0	556	22	1.3
		+40%	508	20	2.7	518	25	2.7	528	30	2.7
		Mean	568	33	8.8	587	29	11.5	606	28	10.4
			Experiment 2								
Size discrimination	sound	−40%	565	30	9.0	565	27	10.0	560	22	7.0
(bigger vs. smaller)		−20%	569	37	14.0	553	19	14.0	553	47	14.0
		−10%	641	38	24.0	593	42	22.0	611	57	31.0
		−5%	605	24	49.0	640	19	44.0	649	36	47.0
		+5%	610	47	16.0	614	38	18.0	647	39	18.0
		+10%	575	29	3.0	535	30	4.0	597	31	13.0
		+20%	521	23	6.0	541	29	5.0	568	53	3.0
		+40%	566	44	2.0	478	26	3.0	491	19	4.0
		Mean	581	34	15.4	565	29	15.0	584	38	17.1
			Experiment 3								
Size discrimination	spoken word	−40%	581	24	8.0	642	44	8.0	630	32	6.0
(bigger vs. smaller)		−20%	649	46	12.0	633	24	9.0	632	32	8.0
		−10%	701	46	15.0	717	44	10.0	713	49	17.0
		−5%	689	32	20.0	786	47	21.0	765	49	23.0
		+5%	738	52	21.0	769	55	30.0	783	44	23.0
		+10%	621	27	8.0	671	27	14.0	672	39	14.0
		+20%	584	31	8.0	608	21	4.0	632	32	5.0
		+40%	574	32	8.0	607	51	5.0	592	25	7.0
		Mean	642	36	12.5	679	39	12.6	677	38	12.9
			Experiment 4								
Size identification	sound	−40%	–	–	–	–	–	–	–	–	–
(same vs. different)		−20%	703	38	17.3	710	44	18.7	738	44	12.7
		−10%	753	41	48.7	775	47	54.7	825	53	53.3
		−5%	711	50	75.3	760	58	74.0	731	37	72.0
		+5%	749	46	66.0	785	60	62.0	745	43	65.3
		+10%	713	41	33.3	747	41	34.7	759	57	44.7
		+20%	671	35	14.7	669	23	9.3	747	47	11.3
		+40%	–	–	–	–	–	–	–	–	–
		Mean	717	42	42.6	741	45	42.2	758	47	43.2

was, however, significant [$F(14,196) = 2.25$, $p < .01$]. Post hoc LSD tests on this interaction revealed significant differences between the congruent and no-sound conditions (with lower error rates reported in the congruent condition) for the +5% and +10% sized variable disks (both ps $< .0001$) and between the congruent and incongruent conditions (with lower error rates reported in the congruent condition) for the +5% sized variable disk ($p < .0001$).[4]

Discussion

The results of Experiment 1 demonstrate that the presentation of an irrelevant sound can modulate the speed with which participants make a speeded classification response regarding the relative sizes of two sequentially presented visual stimuli. The participants responded more rapidly on trials on which a (congruent or incongruent) tone was presented along with the variable-sized disk than on trials on which no sound was presented, presumably reflecting the alerting effect that is typically elicited by the sudden presentation of a sound (see, e.g., Bertelson & Tisseyre, 1969; Posner, 1978; Spence & Driver, 1997).

The most interesting result to emerge from the analysis of Experiment 1 was that the participants responded more rapidly on synesthetically congruent than on synesthetically incongruent sound trials overall. That is, presenting a synesthetically congruent sound simultaneously with the variable-sized disk facilitated the participants' response latencies (mean incongruent − congruent RT difference of 19 msec).[5] A similar but nonsignificant trend was also present in the error data (mean incongruent − congruent differ-
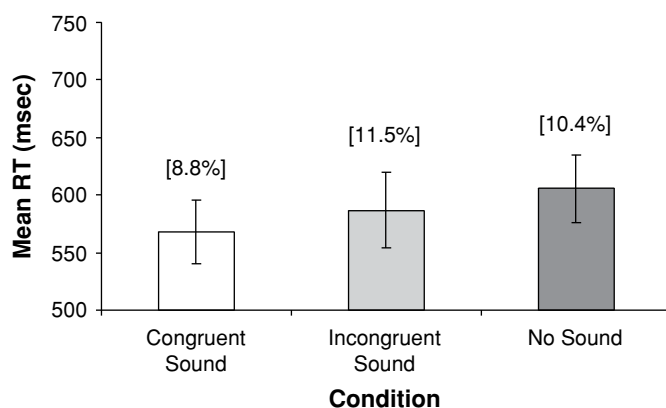
Figure 2. Mean reaction times (RTs, in milliseconds) in the size discrimination (bigger vs. smaller) task in Experiment 1 as a function of condition (congruent sound, incongruent sound, or no sound). The error bars represent the standard errors of the means. The mean percentage of errors for each condition is shown in brackets.

ence in data errors of 2.7%), thus ruling out a speed–accuracy trade-off account of this RT effect (see, e.g., Fitts, 1954; Müller & Findlay, 1987). In other words, the participants in Experiment 1 were able to judge the size of the variable disk relative to that of the standard disk more rapidly (and somewhat more accurately) when the sound that was simultaneously presented was synesthetically congruent with it than when it was synesthetically incongruent. Our results are consistent with previous studies showing synesthetically mediated interactions between auditory and visual stimuli (e.g., Bernstein & Edelstein, 1971; Melara & O'Brien, 1987) and between visual and tactile stimuli (Martino & Marks, 2000). However, our results are the first to demonstrate an auditory synesthetic modulation of judgments of visual stimuli.

It is perhaps worth noting here that the synesthetic association reported in normal individuals in Experiment 1 is consistent with certain multisensory associations reported previously in synesthetic individuals. For example, Zigler (1930) described a form of synesthesia in which the sound of a musical instrument was accompanied by a distinctive 3-D shape. Interestingly, in the 2 synesthetes studied by Zigler, the high-pitched sounds elicited a small, light form whereas the low-pitched sounds elicited a large, dark form (see also Pratt, 1930). This similarity in the correspondences made by both synesthetic and nonsynesthetic individuals had also been reported in several previous studies (e.g., Karwoski & Odbert, 1938; Odbert, Karwoski, & Eckerson, 1942; see Ramachandran & Hubbard, 2001b, for a review).

Although our results show that the presentation of a synesthetically congruent (vs. incongruent) sound can have a significant effect on performance (in terms of reducing RTs on congruent trials in comparison with those on incongruent trials), it is unclear at what stage of information processing this cross-modal interaction takes place. Melara and O'Brien (1987) reported that irrelevant high- and low-frequency sounds failed to influence participants' classification of the vertical position of a dot as either high or low when the sound distractors were presented in sepa-

rate blocks of experimental trials. They argued that when the irrelevant stimulus dimension is held constant across a block of trials, it ceases to exist as a dimension, and, as a consequence, congruency effects should no longer occur. Melara and O'Brien concluded that it is only when the relevant and irrelevant dimensions are both varied on an unpredictable basis that variations in the irrelevant dimension will come to have an effect on participants' judgments on the relevant dimension.

In our second experiment, we therefore presented the high- and low-frequency sounds in separate blocks of experimental trials. Under such conditions, we thought it likely that the pitch of the sound would no longer be classified as either high or low on the basis of its absolute value. If a congruency effect was still reported under these conditions, it would support the suggestion that the relationship between auditory frequency and visual size observed in Experiment 1 acts before any classification of the sound is made, suggesting that synesthetic associations in nonsynesthetic individuals are absolute and not relative (as one might infer from the associations reported in synesthetic individuals; cf. Marks, 1975). If, however, congruency effects were to be eliminated by this blocking of the irrelevant stimulus dimension, then this would support Melara and O'Brien's (1987) claim that synesthetic congruency effects in normal individuals (such as the effect reported in Experiment 1) occur only when there is variation along the irrelevant dimension within a block of trials.

EXPERIMENT 2

Method
Participants. Ten new right-handed participants (5 males and 5 females; mean age, 26 years; range, 22–31 years) took part in Experiments 2 and 3 as paid volunteers. Visual acuity was normal or corrected-to-normal, and all of the participants reported normal auditory perception. The participants received a £5 gift voucher in return for their participation in both Experiments 2 and 3. Each experiment took approximately 20 min to complete.

Apparatus, Materials, Design, and Procedure. The experimental setup and procedure were exactly the same as those in Experiment 1, with the following exceptions: The experiment was composed of three blocks of 80 trials, with short breaks between blocks. In one block of trials, the variably sized visual disk was always presented together with the high-frequency (4500-Hz) sound; in a second block, the low-frequency (300-Hz) sound was always presented at the same time as the variably sized disk; and in a third block of trials, no sound was presented at all. The order of presentation of the blocks of trials was randomized for each participant.

Results

The mean RT data from Experiment 2 (see Table 1 and Figure 3) were submitted to an 8×3 repeated measures ANOVA with size of the variable stimulus ($\pm 5\%$ vs. $\pm 10\%$ vs. $\pm 20\%$ vs. $\pm 40\%$ of the diameter of the standard disk) and sound (congruent vs. incongruent vs. no sound) as the variables. This analysis revealed a significant main effect of size [$F(7,63) = 3.78, p < .005$], with response latencies decreasing as the difference between the sizes of the standard and variable disks increased. The analysis of the RT data failed to reveal any main effect of sound [$F(2,18) = 1.61$] or any interaction between size and sound [$F(14,126) = 1.27$], showing a null effect of congruency when the sounds were presented in separate blocks of trials. In fact, the participants in Experiment 2 responded numerically somewhat more rapidly on incongruent than on congruent trials overall, the reverse or the pattern of results in Experiment 1.

A similar ANOVA was performed on the error data. This revealed a significant main effect of size [$F(7,63) = 10.10, p < .0001$], with the error rate increasing as the difference between the sizes of the standard and variable disks decreased, as was expected. Once again, there was no main effect of sound [$F(2,18) = 2.30, p = .12$] nor any interaction between size and sound [$F(14,126) < 1$].

Discussion

The analysis of Experiment 2 failed to reveal any significant difference in performance between the congruent and incongruent sound conditions. Given that Experiment 2 differed from Experiment 1 solely in that the three sound conditions were presented in separate blocks of trials in Experiment 2 (rather than being randomly presented within every block of trials, as in Experiment 1), one might argue that the presentation of the sound affected the participant's responses regarding the size of a visual stimulus only after it had been classified along a dichotomized dimension (high vs. low in our experiment). Our results therefore add support to Melara and O'Brien's (1987) claim that synesthetic congruency effects in nonsynesthetic individuals are relative (and not absolute), occurring only when there is frequent variation of the stimulus in the irrelevant dimension. However, it is perhaps also worth noting that a null, or reduced, congruency effect was reported by Melara and Algom (2003) in a Stroop color classification task (in which participants had to classify the color in which color names were written) when the color of the word was held constant during the experiment. Therefore, the results of Experiment 2 might also be taken to support the claim that uncertainty/unpredictability (induced by trial-by-trial variation in the case of Experiment 1), rather than the absolute-versus-relative coding of the stimulus properties, may be one of the most important factors in determining congruency effects in both synesthetic congruency and Stroop-like judgment tasks (cf. Melara & Algom, 2003; see also Atchley, Kramer, & Hillstrom, 2000; but see Chastain & Cheal, 2001).

Interestingly, the results of Experiment 2 (like those of Experiment 1) showed no interaction between the size of the visual stimulus and the frequency of the sound. This result may also be taken to suggest that the synesthetic interactions observed in Experiment 1 are more categorical than perceptual in nature. Indeed, whether or not a perceptual synesthetic match is present between the particular sizes of the circle and the specific frequencies of the sounds, one might have expected that at least one of the circle sizes may serendipitously have been perceptually matched with the high or low tone, possibly resulting in a facilitatory effect when presented under blocked condi-
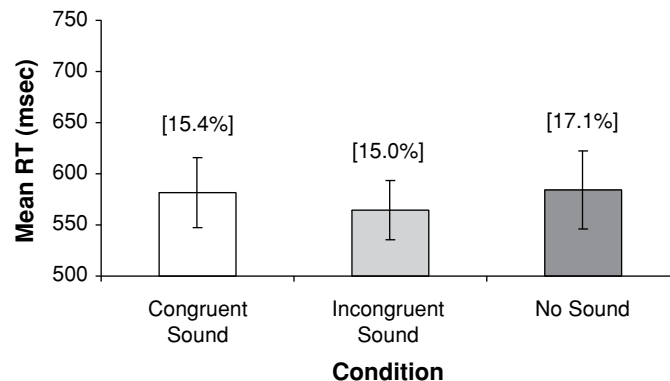

Figure 3. Mean reaction times (RTs, in milliseconds) in the size discrimination (bigger vs. smaller) task in Experiment 2 as a function of condition (congruent sound, incongruent sound, or no sound). The error bars represent the standard errors of the means. The mean percentage of errors for each condition is shown in brackets.

tions. If, for example, the 20% smaller disk happened to provide an exact synesthetic match for the high-frequency sound, a congruency effect might have been expected to have been observed every time that particular disk was paired with that specific sound. However, in order to definitively test for the possible existence of synesthetic perceptual matches between visual stimuli of a particular size and sounds of a particular frequency, participants should first pair visual stimuli of various sizes with sounds of particular frequencies (cf. Walker & Smith, 1984; see also Marks, 2004). Next, the participants would have to judge one dimension of the stimulus (e.g., its visual size) when the other dimension was congruent (in terms of the previously matched stimulus pair) or incongruent (as in pairings of stimuli occurring at random). This procedure was not adopted in the present experiment, which makes it difficult to make any specific claims regarding the perceptual versus categorical nature of the interactions reported here. Further research should therefore be conducted to determine more clearly whether or not perceptual pairings between the specific frequencies of the sounds and the particular sizes of the visual stimuli are present in the cognitive systems of nonsynesthetic individuals.

One possible explanation for the synesthetic interactions reported in nonsynesthetes relates to the role played by the mediation of the verbal codes used to classify stimuli along a given dimension (see Melara & O'Brien, 1987). According to this interpretation, the perceptual information is first converted into the appropriate linguistic code, which is subsequently used to perform the classification task. If the relevant stimulus to be classified and the irrelevant stimulus to be ignored share the same linguistic label (e.g., *high* and *low* are linguistic labels shared by the dimensions of sound elevation and sound frequency; Pratt, 1930), synesthetic interactions (i.e., congruency effects or Garner interference) may arise. This interpretation, however, seems to be disconfirmed as a general account of synesthetic interactions in nonsynesthetes by reports of dimensional interactions, as in the case of color and pitch (see, e.g., Melara & O'Brien, 1987) or brightness and loudness (Marks, 1987), where no obvious sharing of linguistic labels is present between stimuli in the target modality and those in the distractor modality.

An alternative verbal explanation of synesthetic associations states that they arise because of regularities in word usage (see, e.g., Long, 1977). According to this interpretation, attributes that do not share a label (e.g., brightness and loudness) can still produce semantic interactions because they are frequently associated with one another in everyday language. This semantic account of synesthetic interactions predicts that similar results should be obtained whether physical stimuli (i.e., high- and low-frequency sounds) or the words used to classify these stimuli (i.e., the words *high* and *low*) are presented.

In order to explore the role of semantic information in the synesthetic relationship between the pitch of a sound and the size of a stimulus in Experiment 3, we presented spoken words ("high" and "low") rather than pure tones together with the variable-sized disks. If a congruency

effect were to be demonstrated under such conditions, it would provide a clear demonstration that the semantic content of a stimulus can be sufficient to elicit synesthetic interactions, at least under certain conditions.

EXPERIMENT 3

Method

Participants. The same 10 right-handed participants (5 males and 5 females) who had taken part in Experiment 2 took part in Experiment 3. The order of presentation of the two experiments was counterbalanced across participants.

Apparatus, Materials, Design, and Procedure. The experimental setup and procedure were exactly the same as those in Experiment 1, with the following exceptions. The high- and low-frequency sounds were replaced by the spoken words "high" and "low." The fundamental frequencies of the spoken words and the genders of the speakers used in the present experiment were the same for the two words. The presentation of the words lasted for 400 msec. The matching of the fundamental frequency of the spoken words was deemed necessary given empirical observations that the word "high," when spoken normally, is typically perceived as higher in frequency than the word "low" (cf. McClain, 1983).

Results

The mean RT data from Experiment 3 (see Table 1 and Figure 4) were submitted to an 8×3 repeated measures ANOVA with size of the variable stimulus ($\pm 5\%$ vs. $\pm 10\%$ vs. $\pm 20\%$ vs. $\pm 40\%$ of the diameter of the standard disk) and word congruency (congruent vs. incongruent vs. no word) as factors. This analysis revealed a significant main effect of size [$F(7,63) = 10.66, p < .001$], with response latencies decreasing as the difference between the sizes of the standard and variable disks increased. The analysis also revealed a significant main effect of word congruency [$F(2,18) = 8.17, p < .05$]. Post hoc LSD tests revealed a significant difference between the congruent and incongruent word conditions ($p = .003$) and between the congruent and no-word conditions ($p = .003$), but not between the no-word and incongruent word conditions ($p = .87$). The interaction between size and word was not significant [$F(14,126) < 1$].

A similar ANOVA on the error data revealed a significant main effect of size [$F(7,63) = 7.63, p < .0001$], with errors increasing as the difference between the sizes of the standard and variable disks decreased. The participants made similar numbers of errors across the three different levels of the sound factor (see Figure 4), ruling out a speed–accuracy trade-off account of the RT data. None of the other main effects or interactions was significant [$F(2,18) < 1$ and $F(14,126) < 1$ for the main effect of word and the size $\times$ word interaction, respectively].

Discussion

The participants in Experiment 3 responded significantly more rapidly in the congruent than in the incongruent word condition (mean congruency effect of 37 msec). This result demonstrates that synesthetic congruency effects can be mediated by the semantic meaning of a word presented together with the visual stimulus. These results therefore raise the possibility that congruency effects,
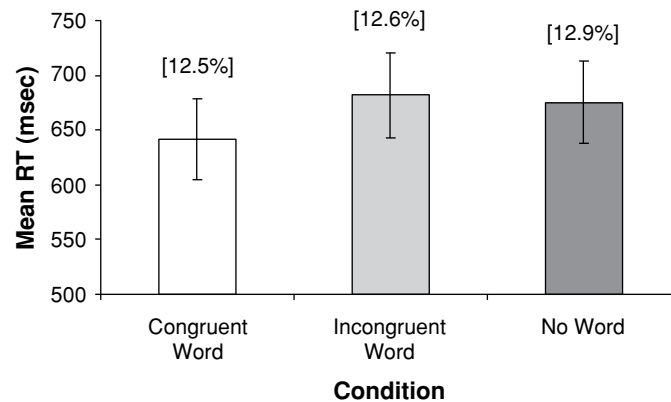
Figure 4. Mean reaction times (RTs, in milliseconds) in the size discrimination (bigger vs. smaller) task in Experiment 3 as a function of condition (congruent word, incongruent word, or no word). The error bars represent the standard errors of the means. The mean percentage of errors for each condition is shown in brackets.

such as those reported in the present study, may arise at a semantic level of information processing (see, e.g., Long, 1977; Martino & Marks, 1999). That is, they may occur at a level at which the sensory modality in which the irrelevant stimulus is presented is in some sense no longer relevant (e.g., at the level of abstract size representation in the present study).

The results of Experiment 3 appear to be in agreement with the findings of Walker and Smith (1984), who reported that RTs to discriminate different adjectives (e.g., *big*, *bright*, *sharp*) were affected by synesthetically incongruent irrelevant auditory and tactile stimuli. By contrast, Pomerantz (1985) failed to demonstrate any effect of the visually presented words *high* and *low* on the speeded classification of the pitch of a sound. The most obvious difference between Pomerantz's (1985) study and the experiment reported here is that the words were presented visually in the former and verbally (i.e., auditorily) in the latter. Only further research will clarify whether or not this is the critical difference between the two experiments.

It is worth noting that the RTs reported in Experiment 3 were longer overall than those obtained in Experiment 2. One possible account for such a difference might relate to the presence of word stimuli in certain trials. That is, even if not explicitly required by the task, prior to responding participants might have analyzed the auditory stimuli presented up to the level of their semantic meanings, thus giving rise to longer processing latencies overall. However, this interpretation can easily be ruled out by the fact that the RTs in Experiment 3 were longer than those in Experiment 2, even under conditions in which no words were presented (i.e., on the neutral trials, which were otherwise identical across the two experiments).

An alternative account of the slowing of responses observed in Experiment 3 relates to the fact that the presentation of word stimuli in the majority of the trials may have engaged a sort of "general verbal strategy" in the participants' processing of the stimuli. According to this

view, both the visual and the auditory stimuli could have been processed in Experiment 2 at a semantic rather than at a perceptual level. Consequently, one might think that synesthetic interactions can occur at different levels of information processing as a function of the specific pattern of stimuli presented (i.e., at a more perceptual level for pure tone stimuli and at a more semantic level for verbal stimuli). This might also be consistent with the view, raised by the Stroop literature, that different patterns of interference between concurrently presented stimuli may depend on the specific cognitive system used to process the information (see Virzi & Egeth, 1984, 1985).

In Experiments 1 and 3, we reported effects of the frequency of sounds and of the words used to classify the sounds on the participants' responses to the size of a visual stimulus. However, given that the perceptual quality of the variable visual stimulus (i.e., bigger or smaller than the standard) and the two alternative responses (i.e., left foot for bigger stimulus and right foot for smaller stimulus) shared the same verbal labels, one might argue that the congruency effect reported in our earlier experiments may have been related to some sort of response priming by the irrelevant sound along the synesthetic dimension (cf. Maddox, 1992; Wang & Proctor, 1996). That is, the interactions between the frequency of a sound and the size of a visual stimulus may have been limited to the stage of information processing at which the response was selected/programmed. In an attempt to rule out this response priming account of our data, we conducted a final experiment in which the response dimension was orthogonal to the dimension in which the visual stimuli varied (cf. Walker & Smith, 1984). Specifically, the participants now had to determine whether the size of the second disk was the same as or different from that of the first disk. In this case, any effect of the sound that was still present could no longer be attributed to a bias in the response selection process, given that the variable-sized disks that were either bigger or smaller than the standard disk were both

mapped onto the same "different" response (see Wang & Proctor, 1996).

EXPERIMENT 4

Method

Participants. Fifteen new right-handed participants (4 males and 11 females; mean age, 24 years; range, 20–31 years) took part in this experiment as paid volunteers. Visual acuity was normal or corrected-to-normal, and all the participants reported normal auditory perception. The experiment took approximately 20 min to complete.

Apparatus, Materials, Design, and Procedure. The experimental setup and procedure were exactly the same as those in Experiment 1, with the following exceptions. The diameter of the variable disk was now ±5%, ±10%, ±20%, or equal to the diameter of the standard disk. Each of the variable disks whose diameter was different from that of the standard disk was presented 10 times for the high-frequency, 10 times for the low-frequency, and 10 times for the no-sound conditions. The variable disk of the same diameter as the standard disk was presented 20 times for each sound condition, giving rise to a total of 240 trials for each participant. The participants were instructed to lift the left foot whenever the size of the second disk appeared to be the same as that of the first disk, and to lift the right foot whenever the size of the second disk appeared to be different from that of the first disk.

Results

For the statistical analyses reported here, the mean RTs and number of errors for each participant were separated into congruent, incongruent, and no-sound conditions for the trials in which the variable disk was different in size from the standard disk (see Table 1). Note that the data from the trials in which the variable and standard disks were of the same size were not analyzed, since they were unclassifiable as either congruent or incongruent.

The mean RT data from Experiment 4 (see Figure 5) were submitted to a 6 × 3 repeated measures ANOVA with size of the variable stimulus (±5% vs. ±10% vs. ±20% of the diameter of the standard disk) and sound (congruent vs. incongruent vs. no sound) as factors. This analysis revealed a significant main effect of size [$F(2,28) = 8.15$, $p = .001$], with a trend toward an increase in RTs as the size of the variable disk approached that of the standard disk. The main effect of sound was also significant [$F(5,70) = 3.34$, $p < .01$]. Post hoc LSD tests revealed a significant difference between the congruent and incongruent sound conditions ($p = .02$) and between the congruent sound and no-sound conditions ($p = .0006$), but failed to reveal a significant difference between the no-sound and incongruent sound conditions ($p = .11$). Once again, the longest RTs were reported in the neutral condition, and the shortest in the congruent condition.

A repeated measures ANOVA on the error data revealed a significant main effect of size [$F(5,70) = 30.76, p < .0001$], with error rates increasing as the difference between the size of the standard and the size of the variable disk decreased. The participants made similar numbers of errors across the three different levels of the sound factor (see Figure 5), ruling out a speed–accuracy trade-off account of the RT data. None of the other main effects or interactions was significant [$F(2,28) < 1$ for the main effect of size and $F(10,140) = 1.22$ for the size × sound interaction].

Discussion

The results of Experiment 4 are consistent with those of Experiments 1 and 3 in that they show both an overall facilitatory effect of the presence of the sound on response latencies (i.e., a warning signal, or alerting, effect—see Posner, 1978; Spence & Driver, 1997) and faster responses on congruent trials than on incongruent trials (mean synesthetic congruency effect of 24 msec in the RT data). Once again, our results show the presence of a synesthetic interaction between the frequency of the sound and relative judgments
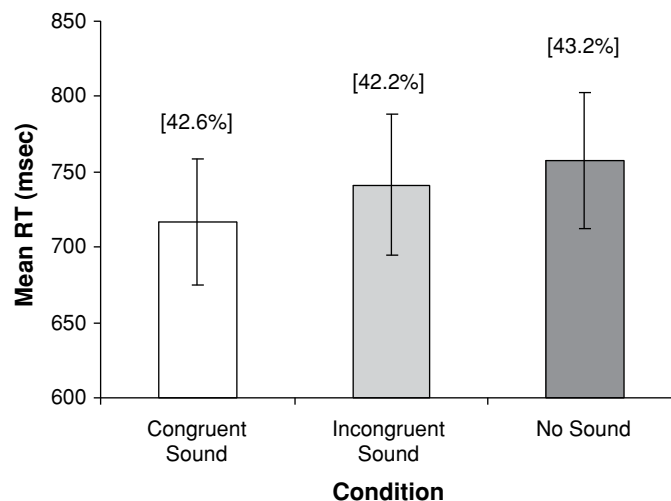

Figure 5. Mean reaction times (RTs, in milliseconds) in the size identification (same vs. different) task in Experiment 4 as a function of condition (congruent sound, incongruent sound, or no sound). The error bars represent the standard errors of the means. The mean percentage of errors for each condition is shown in brackets.

of visual size. The results of Experiment 4 therefore show that the congruency effect reported in Experiment 1 cannot simply be attributed to a response selection bias (see, e.g., Wang & Proctor, 1996), since both the "bigger than the standard" and the "smaller than the standard" response trials were now associated with the same response (i.e., "different from the standard").

GENERAL DISCUSSION

The results of the four experiments reported in the present study confirm the view that people cannot help but process auditory information even when it is irrelevant to their visual task. The RT data show that even when the participants were instructed to ignore the task-irrelevant auditory stimuli, they were influenced by the frequency of the irrelevant sounds (provided that they varied unpredictably on a trial-by-trial rather than a block-by-block basis) when judging the relative sizes of two visually presented stimuli, giving rise to a cross-modal synesthetic congruency effect. Our results therefore suggest that auditory and visual information regarding stimulus size appears to be coded along a synesthetic dimension.

The lack of a difference in the point-of-subjective-equality (PSE) data from Experiment 1 (see note 4) might be taken to indicate that the effect of the irrelevant sound is to speed up participants' responses on the visual size discrimination task, rather than to directly influence the perceived size of the visual disk presented at the same time as the sound. The effect reported here might then be interpreted in terms of some kind of congruency-based prior entry effect (see Shore & Spence, 2005; Spence, Shore, & Klein, 2001). That is, stimuli presented to different sensory modalities may access consciousness more rapidly if they are congruent across a spatial dimension (cf. Driver & Spence, 2004) or, as in the present experiments, across a synesthetic dimension (cf. van Wassenhove, Grant, & Poeppel, 2005).

Walker and Smith (1984) demonstrated that high-pitched sounds and small objects share some cross-modal qualities, as do low-pitched sounds and big objects (e.g., high-pitched sounds and small handled objects are both judged to be sharp, thin, light, fast, and little). In the present study, we demonstrated that there is a sharing of qualities not only between haptic size and auditory frequency, but also between visual size and auditory frequency—in particular, that there is a connection between smaller objects and higher frequency sounds and between bigger objects and lower frequency sounds. The appropriate explanation for the existence of such cross-modal connections needs to be studied more thoroughly in future research, but some clues may emerge from previous studies of synesthetic individuals.

One interpretation for the existence of synesthesia is in terms of the partial failure in the process that eliminates redundant connections between specific sensory brain areas (see Maurer, 1997; Ramachandran & Hubbard, 2001a), whereas others have interpreted synesthetic correspondences as a specific form of learning (e.g., Howells, 1944; Kelly, 1934; Ward & Simner, 2003). The

equivalence between cross-modal dimensions responsible for the congruency effects reported in the present study as well as those reported in previous studies (see Marks, 2004, for a review) might therefore be thought simply to mirror the associations present in nature (see McMahon & Bonner, 1983; Pratt, 1930). Such connections could be either genetically hardwired in the cognitive system or simply learned (given the many cross-modal associations present in nature), thus reflecting adaptations to the cross-modal characteristics of the world. With respect to the latter point, it has been argued that the correspondence between the pitch of sounds and the visual size of stimuli might not necessarily be intrinsic to sensory processing or perception, but rather may have been derived from people's prior experience with the resonance properties of stimuli varying in mass (see Marks, 2000; cf. Brown, 1958; Grassi, 2005; Osgood, Suci, & Tannenbaum, 1957; van den Doel & Pai, 1998).

Synesthetic congruency effects in speeded classification tasks have also been reported in individuals with synesthesia. Synesthetes appear unable to avoid the processing of irrelevant information when it is modulated across their synesthetic correspondences (see, e.g., Dixon, Smilek, Cudahy, & Merikle, 2000; Wollen & Ruggiero, 1983). There appear to be certain correspondences between synesthetic qualities of stimuli in different sensory modalities shared by synesthetic and nonsynesthetic individuals (see, e.g., Odbert et al., 1942). Moreover, these correspondences have been reported to have a constant directionality (i.e., light gray is always judged as corresponding to loud tones and dark gray to soft tones; Stevens & Marks, 1965). One difference between synesthetic and nonsynesthetic individuals is that, whereas for the former group the nature of the synesthetic congruency is idiosyncratic, for normal individuals it appears to be consistent across the population. Another difference is that for synesthetes there is a "perceptual consequent" in one sensory modality attributable to the presentation of the inducing stimulus in the other (or sometimes in the same) sensory modality, which is not reported by nonsynesthetic individuals.

With regard to the results of the present study, one might hypothesize that there is a learned connection between the pitch and the size of objects. Interestingly, people are able to discriminate between objects of different sizes simply by hearing the sounds the objects make when they are dropped onto a surface (see, e.g., Carello, Anderson, & Kunkler-Peck, 1998; Spence & Zampini, 2006). Furthermore, information regarding vocalization pitch is used by certain species of animals to estimate the size of their competitors (see, e.g., Bee, Perrill, & Owen, 2000; Harrington, 1987).[6] However, further developmental research will be needed to determine whether synesthetic associations between the pitches of sounds and the sizes of objects are phylogenetically or ontogenetically determined (see Marks, Hammeal, & Bornstein, 1987). Recent results obtained with 30- to 36-month-old children (Mondloch & Maurer, 2004) have revealed the presence of a synesthetic association between the sizes of balls and the pitches of sounds (i.e., higher pitched sounds were associated with

smaller balls, and lower pitched sounds with larger balls). Such results may be taken to suggest a hardwired mechanism underlying synesthetic associations. Moreover, what appears clear from these observations is the presence of a link between sensory modalities tuned to certain natural associations (see R. Walker, 1987). It has been argued that adaptive or "hyperlearned" associations between different sensory modalities may group together better (in a gestalt sense; see, e.g., Pomerantz, 1981), thus resulting in faster and/or more efficient responding. Therefore, some synesthetic associations should probably be considered in terms of adaptive behavior rather than solely in terms of a failure of selective attention (see Martino & Marks, 2000).

The results of Experiment 2 highlight the fact that synesthetic associations in nonsynesthetic individuals are relative rather than absolute, and that congruency effects occur only when there is variation along the irrelevant dimension on a trial-by-trial basis (at least for the high- and low-frequency tones used in the present study). This result suggests that the connection between the frequency of a sound and the size of a visual stimulus reported in the present study occurs after the perceptual attributes of the stimuli have been classified cognitively. These data contribute to moving the locus of synesthetic interactions from a purely perceptual to a higher level of information processing. Support for this suggestion comes from the results of Experiment 3, in which an effect on the classification of the sizes of the visual stimuli was demonstrated using the spoken words "high" and "low" rather than the high- and low-frequency auditory stimuli of Experiments 1, 2, and 4. Nevertheless, it is worth noting that the interaction between spoken words and the sizes of visual stimuli reported in Experiment 3, in which there was no sharing of the verbal labels used to classify the two dimensions of the stimuli (e.g., *bigger/smaller* for the visual stimuli and *higher/lower* for the frequencies of the sounds), argues against a purely verbal interpretation of the synesthetic associations.

Long (1977) suggested that synesthetic associations may arise because of the regularity of word usage. Perceptual attributes, whether they share verbal labels or not, can produce interactions because of their frequent association in natural language. Nevertheless, the suggestion that attributes that are frequently associated in natural language are responsible for synesthetic interactions can also be viewed in the opposite way. That is, certain linguistic terms may be frequently associated in language because they share synesthetic qualities. For instance, it has been demonstrated that when participants have to name meaningless shapes, they exhibit a consistent tendency to associate pictures composed by sharp contours with pseudowords containing sharp phonemic inflections (e.g., *takete*) and pictures composed by rounded contours with pseudowords containing rounded phonemic inflections (e.g., *baluma*; see Köhler, 1929). On the basis of such observations, Ramachandran and Hubbard (2001b) proposed that language might actually originate from synesthesia, driven by the natural constraints on the ways in which sounds are mapped to objects, rather than the other way around. Such speculation clearly warrants further empirical research.

The results of the present study suggest that the cognitive basis of synesthetic associations in nonsynesthetic individuals neither is purely perceptual (as is suggested by the results of the PSE analysis in Experiment 1 and by the results of Experiment 2; see also note 5) nor simply reflects response priming (Experiment 4) or the consequence of the sharing of verbal labels used to classify the stimuli (Experiments 1 and 3). Synesthetic associations might be based on a presemantic/semantic level of processing at the stage in which an abstract, amodal representation of the stimuli (shaped on the basis of natural constraints) might be set up (cf. van Wassenhove et al., 2005).

Whatever the correct interpretation of our results turns out to be, they nevertheless provide the first empirical evidence for a synesthetic modulation of participants' responses to the size of visual stimuli by means of variations in the frequency of simultaneously presented sounds. Our results are consistent with previous findings of cross-modal synesthetic correspondences between the elevation of visual stimuli and sound frequency (Melara & O'Brien, 1987) and between visual lightness and tactile frequency (Martino & Marks, 2000).

REFERENCES

ATCHLEY, P., KRAMER, A. F., & HILLSTROM, A. P. (2000). Contingent capture for onsets and offsets: Attentional set for perceptual transients. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 594-606.

BARON-COHEN, S., HARRISON, J., GOLDSTEIN, L. H., & WYKE, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception*, **22**, 419-426.

BEE, M. A., PERRILL, S. A., & OWEN, P. C. (2000). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology*, **11**, 169-177.

BERNSTEIN, I. H., & EDELSTEIN, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, **87**, 241-247.

BERTELSON, P., & TISSEYRE, F. (1969). The time-course of preparation: Confirmatory results with visual and auditory warning signals. *Acta Psychologica*, **30**, 145-154.

BROWN, R. (1958). *Words and things*. Glencoe, IL: Free Press.

CARELLO, C., ANDERSON, K. L., & KUNKLER-PECK, A. J. (1998). Perception of object length by sound. *Psychological Science*, **9**, 211-214.

CHASTAIN, G., & CHEAL, M. (2001). Attentional capture with various distractor and target types. *Perception & Psychophysics*, **63**, 979-990.

CLARK, H. H., & BROWNELL, H. H. (1975). Position, direction, and their variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 339-352.

DIXON, M. J., SMILEK, D., CUDAHY, C., & MERIKLE, P. M. (2000). Five plus two equals yellow. *Nature*, **406**, 365.

DRIVER, J., & SPENCE, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179-220). Oxford: Oxford University Press.

FINNEY, D. J. (1964). *Probit analysis: Statistical treatment of the sigmoid response curve*. London: Cambridge University Press.

FITCH, W. T., & REBY, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London: Series B*, **268**, 1669-1675.

FITTS, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, **47**, 381-391.

GARNER, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

GARNER, W. R. (1977). The effect of absolute size on the separability of

the dimensions of size and brightness. *Bulletin of the Psychonomic Society*, **9**, 380-382.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception & Psychophysics*, **67**, 274-284.

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences*, **5**, 36-41.

Harrington, F. H. (1987). Aggressive howling in wolves. *Animal Behaviour*, **35**, 7-12.

Howells, T. (1944). The experimental development of color–tone synesthesia. *Journal of Experimental Psychology*, **34**, 87-103.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, **96**, 29-44.

Karwoski, T. F., & Odbert, H. S. (1938). Color-music. *Psychological Monographs*, **50**(2, Whole No. 22).

Kelly, E. (1934). An experimental attempt to produce artificial chromaesthesia by the technique of the conditioned response. *Journal of Experimental Psychology*, **17**, 315-341.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Long, J. (1977). Contextual assimilation and its effect on the division of attention between nonverbal signals. *Quarterly Journal of Experimental Psychology*, **29**, 397-414.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147-180). Hillsdale, NJ: Erlbaum.

Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, **82**, 303-331.

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

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 384-394.

Marks, L. E. (2000). Synesthesia. In E. Cardeña, S. J. Lynn, & S. C. Krippner (Eds.), *Varieties of anomalous experience: Examining the scientific evidence* (pp. 121-149). Washington, DC: American Psychological Association.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 85-105). Cambridge, MA: MIT Press.

Marks, L. E., Hammeal, R. J., & Bornstein, M. H. (1987). Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development*, **52**, 1-102.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Test of the semantic coding hypothesis. *Perception*, **28**, 903-923.

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception*, **29**, 745-754.

Maurer, D. (1997). Neonatal synaesthesia: Implications for the processing of speech and faces. In S. Baron-Cohen and J. Harrison (Eds.), *Synaesthesia: Classic and contemporary readings* (pp. 224-242). Oxford: Blackwell.

McClain, L. (1983). Stimulus–response compatibility affects auditory Stroop interference. *Perception & Psychophysics*, **33**, 266-270.

McMahon, T. A., & Bonner, J. T. (1983). *On size and life*. New York: Scientific American.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review*, **110**, 422-471.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, **116**, 323-336.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, **4**, 133-136.

Müller, H. J., & Findlay, J. M. (1987). Sensitivity and criterion effects in the spatial cuing of visual attention. *Perception & Psychophysics*, **42**, 383-399.

Odbert, H. S., Karwoski, T. F., & Eckerson, A. B. (1942). Studies in synesthetic thinking: I. Musical and verbal associations of color and mood. *Journal of General Psychology*, **26**, 153-173.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 755-775.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141-180). Hillsdale, NJ: Erlbaum.

Pomerantz, J. R. (1985, November). *Selective attention: Difference between Stroop and Garner interference*. Paper presented at the 26th Annual Meeting of the Psychonomic Society, Boston.

Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 53-89). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, **13**, 278-285.

Ramachandran, V. S., & Hubbard, E. M. (2001a). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London: Series B*, **268**, 979-983.

Ramachandran, V. S., & Hubbard, E. M. (2001b). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, **8**, 3-34.

Rubinstein, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica*, **111**, 141-154.

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 515-537.

Shore, D. I., & Spence, C. (2005). Prior entry. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 89-95). Amsterdam: Elsevier, North-Holland.

Spence, C., & Driver, J. (1997). Audiovisual links in endogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.

Spence, C., & McDonald, J. (2004). The crossmodal consequence of the exogenous spatial orienting of attention. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 3-25). Cambridge, MA: MIT Press.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, **130**, 799-832.

Spence, C., & Zampini, M. (2006). Auditory contributions to multisensory product perception. *Acta Acustica United With Acustica*, **92**, 1009-1025.

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, **54**, 407-411.

Stumpf, K. (1883). *Tonpsychologie I* [Psychology of tone]. Leipzig: Hirzel.

Taylor, R. L., & Campbell, G. T. (1976). Sensory interaction: Vision is modulated by hearing. *Perception*, **5**, 467-477.

van den Doel, K., & Pai, D. K. (1998). The sounds of physical shapes. *Presence*, **7**, 382-395.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, **102**, 1181-1186.

Virzi, R. A., & Egeth, H. E. (1984). Is meaning implicated in illusory conjunctions? *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 573-580.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, **13**, 304-319.

Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, **13**, 75-81.

Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics*, **42**, 491-502.

Wang, H., & Proctor, R. W. (1996). Stimulus–response compatibility as a function of stimulus code and response modality. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1201-1217.

Ward, J., & Simner, J. (2003). Lexical–gustatory synaesthesia: Linguistic and conceptual factors. *Cognition*, **89**, 237-261.

Wollen, K. A., & Ruggiero, F. T. (1983). Colored-letter synesthesia. *Journal of Mental Imagery*, **7**, 83-86.

Zigler, M. J. (1930). Tone shapes: A novel type of synaesthesia. *Journal of General Psychology*, **3**, 276-287.

NOTES

1. The effects reported by Garner (1974) are related to the tasks used to classify the properties of the stimuli—in particular, the baseline task, in which the value on the dimension to be classified is varied whereas the value on the irrelevant dimension is held constant; and the orthogonal task, in which the values on both dimensions are varied randomly on a trial-by-trial basis. Poorer performance in the orthogonal than in the baseline task is referred to as *Garner interference*. *Congruence effects* refer to the processing of stimuli that have a directional dimension (e.g., direction of movement, higher vs. lower elevation, or brighter vs. darker color). Congruent stimuli share the same directional value, whereas incongruent stimuli have opposite directional values. Congruence effects account for the improved performance normally seen on congruent trials in comparison with incongruent trials in the orthogonal task.

2. The word *synesthesia* (from the Greek *syn* [together] + *aesthesis* [perception]) has been used to describe "a conscious experience of systematically induced sensory attributes that are not experienced by most people under comparable conditions" (Grossenbacher & Lovelace, 2001, p. 36). Synesthetes are individuals who, when presented with a specific stimulus, report an additional sensory experience (in either the same or a different sensory modality) that is not experienced by nonsynesthetes. For example, synesthetes with "colored hearing" report seeing a specific color when they hear a certain word spoken (see, e.g., Marks, 1975). Such synesthetic experiences appear consistent over time (Baron-Cohen, Harrison, Goldstein, & Wyke, 1993) and are often elicited automatically (Dixon, Smilek, Cudahy, & Merikle, 2000). Such findings have led some researchers to suggest that the correspondences made by both synesthetic and nonsynesthetic individuals might therefore be based on the same underlying perceptual mechanisms (see, e.g., Karwoski & Odbert, 1938).

3. In their preliminary study, Walker and Smith (1984) asked participants to haptically explore wooden knobs of different sizes and to rate them on 7-point scales with anchors such as *little–big*, *thin–thick*, and *light–heavy*. The same rating procedure was used for judging the quality of pure tones that varied in frequency. The correspondences found in their study (e.g., that both a high-pitched sound and a small, handled object were judged to be sharp, thin, light, fast, and little) were then used in a second experiment involving a variant of the Stroop interference task.

4. In order to further analyze the effect of the congruency of the irrelevant sound on the perceived size of the variable disk, psychophysical functions were calculated for each sound condition. For each participant, the probability of judging each variable-sized disk as bigger than the standard disk was transformed into a corresponding *z*-score (see, e.g., Finney, 1964). The regression analyses for the functions fitting the data revealed a significant linear relationship between the size of the variable stimulus and the corresponding *z*-score for all of the sound conditions [$F(1,6) = 30.2$, $r^2 = .834$, $p < .005$ for the high-frequency sound condition; $F(1,6) = 28.4$, $r^2 = .826$, $p < .005$ for the low-frequency sound condition; $F(1,6) = 28.34$, $r^2 = .825$, $p < .005$ for the no-sound condition]. The slopes and intercepts of the fitted functions were then calculated, and the point at which the functions crossed the point of subjective equality (PSE—i.e., the point at which the participants judged each of the two disks as being larger on an equal number of trials) was determined for each condition. The PSE data were then submitted to a one-way repeated measures ANOVA with the variable of sound, which revealed no significant differences among the three conditions [$F(2,28) = 1.06$, $p = .36$]. The analysis of the PSE data therefore suggests that the presentation of the sound had no effect on the perceived size of the disks, despite its having a significant effect on response latencies.

5. The fact that a neutral sound condition was not included in Experiment 1 makes it rather difficult to make any strong claims concerning whether the results reflect facilitation on congruent trials or interference on incongruent trials. However, the RT data obtained in Experiment 2 can be used to provide an approximate value for a neutral sound condition. A comparison of these values with the RTs obtained in Experiment 1 (cf. Sabri, Melara, & Algom, 2001) shows that synesthetically congruent trials led to a 13-msec facilitation effect, whereas synesthetically incongruent trials led to a 22-msec interference effect. Unfortunately, however, the fact that different participants took part in the two experiments means that no strong conclusions can be drawn on the basis of this comparison.

In order to further assess whether the results of Experiment 1 reflect a congruency effect or an interference effect of the sound on responses to the visual stimulus, a neutral white noise condition was used in a control experiment ($N = 12$ participants). The apparatus, materials, design, and procedures were identical to those used in Experiment 1, with the sole exception that the no-sound trials were substituted with neutral noise trials in which white noise of the same duration as the high- and low-frequency sounds (i.e., 300 msec) was used. Under these experimental conditions, no significant effect of the sound on the speed of the participants' responses was observed (although the trends in the RTs and the error data rate were in the expected direction—i.e., the shortest RTs [$M = 592$ msec] and the lowest error rates [$M = 12.6\%$] were reported in the congruent condition and the longest RTs [$M = 599$ msec] and the highest error rates [$M = 14.6\%$] in the incongruent condition, with intermediate performance reported on the neutral white noise trials [$M = 594$ msec and $M = 13.9\%$ errors]). This difference between performance on congruent and incongruent trials would appear too small to allow any significant differences to be demonstrated between the congruent/incongruent conditions and the newly introduced neutral condition. Nevertheless, it is perhaps also worth noting that other researchers have argued that finding an appropriate neutral cue condition is often somewhat problematic in attentional research (e.g., Jonides & Mack, 1984). Indeed, although not informative about target properties, any neutral condition may make additional processing demands and may not necessarily elicit neutral behavioral strategies on the part of participants.

A further analysis of the data obtained in this control experiment was performed in order to determine whether or not the failure to demonstrate a significant effect might be related to sequence effects. We analyzed the difference between congruent and incongruent trials as a function of whether the auditory stimulus on the preceding trial was of the same frequency, different in frequency, or white noise. This analysis did not reveal any significant effect of the previously presented sound on the congruency effect [i.e., on the difference in RTs between congruent and incongruent trials; $F(2,22) = 2.25$, $F(1,11) < 1$, and $F(2,22) < 1$ for the effects of congruency, previously presented sound, and their interaction, respectively].

The pattern of results obtained in this control experiment may also suggest that the introduction of a third sound might have had a detrimental effect on the dichotomy between the high- and low-frequency sounds (by means of which the participants' judgments were presumably being influenced). This would support the view that the effects reported in Experiment 1 were operating more at a semantic level of processing than at a perceptual one. The results obtained, together with the anecdotal observation by participants indicating that the white noise was subjectively classified as having a relatively high frequency in the control study (although physically composed of all the frequencies of the audible spectrum), makes it difficult to draw any firm conclusions regarding the role played by facilitation versus interference over the participants' responses. This important topic deserves further study in future research. Specifically, a within-participants comparison between trials on which the sounds are blocked and trials on which the sounds are presented randomly on a trial-by-trial basis might be made, as an alternative to the use of the white noise neutral condition reported here.

6. Interestingly, a peculiar human adult structure—the descended larynx (thought to represent only an adaptation for articulate speech)—seems to have evolved in other species as a means of influencing perceived body size by decreasing the resonant frequencies of the vocal tract (Fitch & Reby, 2001).