Multisensory temporal order judgments: When two locations are better than one

CHARLES SPENCE
University of Oxford, Oxford, England

ROLAND BADDELEY
University of Sussex, Brighton, England

MASSIMILIANO ZAMPINI and ROBERT JAMES
University of Oxford, Oxford, England

and

DAVID I. SHORE
McMaster University, Hamilton, Ontario, Canada

In Experiment 1, participants were presented with pairs of stimuli (one visual and the other tactile) from the left and/or right of fixation at varying stimulus onset asynchronies and were required to make unspeeded temporal order judgments (TOJs) regarding which modality was presented first. When the participants adopted an uncrossed-hands posture, just noticeable differences (JNDs) were lower (i.e., multisensory TOJs were more precise) when stimuli were presented from different positions, rather than from the same position. This spatial redundancy benefit was reduced when the participants adopted a crossed-hands posture, suggesting a failure to remap visuotactile space appropriately. In Experiment 2, JNDs were also lower when pairs of auditory and visual stimuli were presented from different positions, rather than from the same position. Taken together, these results demonstrate that people can use redundant spatial cues to facilitate their performance on multisensory TOJ tasks and suggest that previous studies may have systematically overestimated the precision with which people can make such judgments. These results highlight the intimate link between spatial and temporal factors in determining our perception of the multimodal objects and events in the world around us.

One of the oldest and most important questions in multisensory information-processing research relates to the extent to which people can discriminate the temporal ordering of stimuli presented in different sensory modalities (e.g., Bloch, 1887; Exner, 1875; Hamlin, 1895; Whipple, Sanford, & Colgrove, 1899; see Smith, 1933, for an early review). The ability to discriminate whether two events are simultaneous or not is particularly important given that synchrony provides one of the most salient cues (along with spatial location) regarding whether stimuli in different sensory modalities should be bound together to form a single multisensory perceptual object or else treated as separate and independent perceptual objects/events (cf. Driver & Spence, 2000; Robertson et al., 2001; Slutsky & Recanzone, 2001; Stein & Meredith, 1993).

A majority of researchers interested in this topic have used the temporal order judgment (TOJ) task to assess people's sensitivity to temporal asynchrony between stimuli presented in different sensory modalities.[1] In a typical TOJ study, participants are presented with two stimuli at varying stimulus onset asynchronies (SOAs) and are required to judge which stimulus was presented first. In their seminal multisensory TOJ study, Hirsh and Sherrick (1961) reported that the interval between stimuli required for participants to judge which modality came first on 75% of the trials (the so-called just noticeable difference, or JND) was 20 msec when only two stimuli were presented. Surprisingly, the JND remained relatively constant across various intramodal and multisensory combinations of auditory, visual, and tactile stimuli. Similar results have been reported in a number of other multisensory TOJ studies (e.g., Bald, Berrien, Price, & Sprague, 1942; Rutschmann & Link, 1964; although see Jaśkowski, Jaroszyk, & Hojan-Jerierska, 1990).[2]

Spatial Redundancy and Multisensory TOJs

Hirsh and Sherrick (1961) may, however, have systematically overestimated people's ability to make multisensory TOJs, because the stimuli in the different sen-

sory modalities were always presented from different spatial locations in their study: Auditory stimuli were presented via an earphone, tactile stimuli were presented to the index finger placed by the participant's side, and visual stimuli were presented from a screen placed directly in front of the participant. Consequently, the participants may simply have responded on the basis of which *location* was stimulated first, rather than on the basis of which *modality* they perceived first (as they had been instructed by Hirsh & Sherrick, 1961; cf. Spence & Driver, 1997). That is, the participants may have used the spatial redundancy to improve their multisensory TOJ performance.[3]

Support for the claim that people may use redundant cues to facilitate their performance on the multisensory TOJ task comes from a study by Taylor (1978), in which redundant tactile cues were shown to facilitate performance on an intramodal tactile TOJ task. Similarly, Hirsh and Sherrick (1961, p. 428) also reported a tendency for redundant pitch cues to facilitate intramodal auditory spatial TOJs. To the best of our knowledge, stimuli in different modalities were presented from different spatial locations in almost all previous multisensory TOJ studies (e.g., Bald et al., 1942; Dinnerstein & Blitz, 1964; Hamlin, 1895; Jaśkowski et al., 1990; Rutschmann & Link, 1964; Smith, 1933; Teatini, Farnè, Verzella, & Berruecos, 1976; Whipple et al., 1899; although see Gengel & Hirsh, 1970, for a solitary exception to this claim). Consequently, it is possible that people may actually be much worse at judging the correct ordering of stimuli presented to different sensory modalities than the majority of the empirical literature would suggest.

The primary aim of the present study was, therefore, to examine whether people can use spatial cues when making multisensory TOJs regarding which modality has been presented first. In Experiment 1, we presented pairs of visual and tactile stimuli from either the same or different spatial locations. The stimuli were separated by an SOA of between 10 and 200 msec, using the method of constant stimuli (see Coren, Ward, & Enns, 1999; Shore, Spence, & Klein, 2001). The participants made unspeeded footpedal discrimination responses regarding *which modality came first*? Our prediction was that if participants can use redundant spatial information when judging which modality was presented first (or else if they respond, when possible, on the basis of *which location was presented first*), JNDs should be larger (i.e., performance should be less accurate) on trials in which the two stimuli were presented from the same spatial location than when they were presented from different locations. In Experiment 2, we used a similar paradigm to investigate the role of spatial redundancy in audiovisual TOJs.

Posture Change and TOJs

Recently, several researchers have reported that changes in posture can dramatically affect tactile temporal perception (e.g., Shore, Spry, & Spence, 2002a, 2002b; Yamamoto & Kitazawa, 2001a, 2001b). For example, Shore

et al. (2002b) showed that tactile spatial TOJs (*left vs. right stimulus first?*) for pairs of identical vibrotactile stimuli presented one to either hand are more accurate (i.e., JNDs are lower) when the hands are placed far apart than when placed close together, despite the fact that the tactile stimulation remained constant across posture change. Similarly, Shore et al. (2002a) and Yamamoto and Kitazawa (2001a) have shown that people's ability to discriminate which hand was stimulated first by a tactile stimulus can be disrupted by simply crossing the hands. The participants in Shore et al.'s (2002a) study achieved JNDs of 34 msec when the hands were positioned in an uncrossed posture and 124 msec when the hands were crossed (so that the right hand lay in the left hemispace and the left hand in the right hemispace) in complete darkness. By contrast, no such spatiotemporal processing deficit was reported for pairs of visual stimuli placed in identical positions to the two tactile stimuli when the participants crossed their hands (Shore et al., 2002a; Yamamoto & Kitazawa, 2001a).

These results, showing an apparent failure to localize tactile stimuli appropriately when the hands are crossed, contrast with the pattern of results now emerging from other research in which people appear to localize touch to the crossed hands without any problem. For example, Kennett and colleagues (e.g., Kennett, Eimer, Spence, & Driver, 2001; Kennett, Spence, & Driver, 2002) have shown that a spatially nonpredictive vibrotactile stimulus (similar to that used by Shore et al., 2002a) presented to either the left or the right hand leads to an automatic (i.e., exogenous) shift of visual spatial attention to the appropriate external location of the vibrotactile stimulus, no matter whether the hands are placed in an uncrossed or a crossed posture. Kennett et al.'s results therefore show that people are capable of localizing vibrotactile stimuli presented to the crossed hands under at least certain conditions (see also Anzola, Bertoloni, Buchtel, & Rizzolatti, 1977, and Bradshaw, Howard, Pierson, Phillips, & Bradshaw, 1992, for similar results).

The question therefore arises as to what is (are) the critical difference(s) between the TOJ studies reported by Shore et al. (2002a) and Yamamoto and Kitazawa (2001a, 2001b) and the spatial cuing studies reported by Kennett et al. (2001; Kennett et al., 2002) and others that account for this difference. It is possible that the difficulty in the TOJ studies stemmed from the fact that people find it difficult to process (or spatially remap) two or more vibrotactile stimuli if they are presented in rapid succession in the crossed posture. This might be due to some form of computational overload or spatiotemporal ambiguity specific to the tactile system (see Shore et al., 2002a, and Weiland, 1960, on this point). Alternatively, the confusion might be caused by differences between the TOJ and the spatial discrimination tasks themselves. For example, participants had to localize the vibrotactile stimuli in order to respond correctly in the TOJ studies, whereas they only had to respond explicitly to visual stimuli (and not to the spatially nonpredictive vibrotactile cue) in Ken-

nett et al.'s (2001; Kennett et al., 2002) spatial cuing studies. However, it seems unlikely that this last explanation will account for the difference, given that several other studies have shown that people can easily localize single tactile stimuli when their hands are crossed over the midline (e.g., Anzola et al., 1977; Bradshaw et al., 1992).

The second major aim of Experiment 1 was, therefore, to assess whether people would find it difficult to localize a vibrotactile stimulus when it was presented together with a visual stimulus in the multisensory TOJ setting. If participants have no difficulty in localizing the vibrotactile stimulus in the crossed posture, we would predict that their performance should be no different from that seen in the uncrossed posture. If, however, people find it difficult to localize a vibrotactile stimulus in the crossed posture when it is presented close in time to a visual event near one or the other hand, we would expect performance to change. More specifically, we would expect that the spatial redundancy gains predicted (and found) in the uncrossed posture when vision and touch are presented from different spatial positions (as compared with the same location) would be eliminated in those blocks of trials in which the participants adopted a crossed-hands posture.

EXPERIMENT 1

Methods

Participants. Ten participants (mean age of 20 years) were recruited to take part in the experiment, which took approximately 60 min to complete. All the participants were naive as to the purpose of the experiment, and none had taken part in any psychophysical experiments before.

Apparatus and Materials. The experiment was conducted in a completely dark sound-attenuated room. The participants were seated at a table facing straight ahead, with their heads resting on a chinrest. A red light-emitting diode (LED; luminance of 64.3 cd/m^2) was placed 62 cm in front of the participants on the table and served as a fixation point. Two identical red LEDs were positioned 26 cm to either side of the fixation light at the same distance from the participant. Two Oticon-A (100Ω) bone conduction vibrators, with vibrating surfaces 1.6 cm in width and 2.4 cm in length, were used to present the vibrotactile stimuli. One vibrator was placed directly in front of each of the two peripheral LEDs (i.e., from the same spatial locations). The vibrators were driven by a 200-Hz sine wave signal from a signal generator. When adopting the uncrossed posture, the participants rested their hands on the table directly in front of them, with their left index fingers resting over the left vibrotactile stimulator and their right index fingers over the right vibrator. When adopting the crossed posture, the participants crossed their hands over the midline so that the left index fingers now rested over the right vibrotactile stimulator and their right index fingers rested over the left vibrator. Visual stimuli consisted of the onset of one or the other target LED for 15 msec, and the tactile stimuli consisted of the onset of either vibrotactile stimulator for 15 msec. (Note that these stimulus durations are somewhat longer that those used in Hirsh & Sherrick's, 1961, seminal study.) Although the visual and tactile stimuli in our study were presented at a clearly suprathreshold level, no specific attempt was made to try and match their intensities.[4] White noise was presented continuously at 75 dB (A) binaurally over headphones throughout the experimental blocks to mask any sound made by the operation of the vibrators.

The participants normally kept two footpedals depressed, one under the toe and the other under the heel of their right foot. The

participants lifted their toes to indicate that the visual stimulus had been presented first and their heels to indicate that the tactile stimulus appeared to have been presented first. Given the use of briefly presented discrete tactile and visual stimuli, the participants presumably could use information from the relative perceived time of onset of the two stimuli, information from the relative perceived offset of the stimuli, or both types of information in making their TOJ responses in the present experiment (cf. Stone et al., 2001). No specific attempt was made to counterbalance the toe–heel response mapping. However, we think it unlikely that this affected performance, given that the participants were required to make *unspeeded* TOJ responses in our study. We used footpedal responses, rather than buttonpress responses, to limit any stimulus–response compatibility effects, given lateral displacement of the hands (see, e.g., Simon, 1990), and also to rule out any problems in the crossed posture condition that might be associated with being uncertain as to which button to press (see Shore et al., 2002a). We have used footpedal responding in a number of our previous multisensory studies (e.g., Spence, Nicholls, Gillespie, & Driver, 1998; Spence, Pavani, & Driver, 2000; Spence, Shore, & Klein, 2001), and participants find it easy to use such a method to respond.

The response made by the participant was indicated by the illumination of one of two LEDs placed directly above and below the fixation light immediately after a response was detected. The upper light was illuminated to indicate a *vision-first* response, and the lower light was illuminated to indicate a *touch-first* response. The inclusion of the feedback lights was designed to facilitate the acquisition of the footpedal response mapping by the participants without giving them any corrective error feedback (which might have affected their performance and, possibly, even their point of subjective simultaneity [PSS]; see Gengel & Hirsh, 1970). Presentation of the stimuli and monitoring of the responses was controlled by an IBM 486 compatible microcomputer using a program written in Turbo Pascal 6.0. Timing was controlled by an 82C54 interval-timer chip on one of the input–output cards (DCM-16 and DOP-24; Blue Chip Technology, Deeside, U.K.) that interfaced to the LEDs, vibrotactile stimulators, and response pedals.

Design. There were 40 equiprobable conditions produced by crossing the three within-subjects factors: stimulus location (same vs. different position), SOA between target stimuli (−200, −90, −55, −30, −10, +10, +30, +55, +90, and +200 msec; negative SOAs indicate that the visual stimulus was presented first, whereas positive values indicate that the tactile stimulus was presented first), and posture (uncrossed- vs. crossed-hands posture). Our decision to use a relatively broad range of SOAs (which were much wider that those used by Hirsh & Sherrick, 1961) was based on our previous work on prior entry, using a similar TOJ paradigm (see Shore et al., 2001; Spence, Shore, & Klein, 2001). The participants completed two blocks of 20 practice trials, followed by eight blocks of 80 test trials. The participants performed each block of trials in either the crossed or the uncrossed posture, with posture alternating between successive blocks of trials. Half of the participants started with their hands in the uncrossed posture, whereas the remainder started with their hands in the crossed posture. The participants were presented with 64 single-target trials before the start of the experimental session to familiarize them with the stimulus–response mapping used in this experiment. The SOAs were doubled in the first practice block of TOJ trials to facilitate the participant's acquisition of the TOJ discrimination task. The visual and tactile stimuli were equally likely to be presented from the left as from the right of fixation.

Procedure. At the start of the experiment, the participants were informed that they would be presented with one visual and one tactile stimulus on each trial and that each of these stimuli could be presented randomly from either the left or the right on each trial. They were informed that they would have to decide on each trial which stimulus modality had been presented first and that they would probably find this discrimination difficult on some occa-

sions. The participants were also informed that the task was unspeeded and that they should respond only when confident of their response. The fixation light was illuminated at the beginning of each trial. The participants were instructed prior to the experiment to maintain fixation on this central red LED throughout each block of trials. After a delay of 750 msec, the first stimulus was presented from either the left or the right of fixation.[5] The second stimulus was presented after the SOA specified for that condition. If the participants made an anticipatory response (i.e., prior to the onset of the first stimulus) or failed to make a response before the trial was terminated (3,000 msec after the onset of the first stimulus), an error feedback signal was presented (<1% of the trials). This consisted of the flickering of the fixation light for 1,000 msec. Otherwise, the participant's response was indicated by the illumination of one of the central lights for 500 msec. The fixation light was illuminated to indicate the start of the next trial 750 msec after the end of the preceding trial.

Results

The population average response for the uncrossed and crossed postures of Experiment 1 are shown in Figures 1A and 1B. Although this is an intuitive way of viewing the data, it can be statistically misleading, since if the PSS varied between participants, then in the group average this would misleadingly appear as a change in slope (see also Estes, 1956). The calculation of both JNDs and slopes is far more accurate if a model is fit to each participant's data and, then, the parameters extracted from the resulting curve. This procedure of estimating parameters from a fitted psychometric curve is a standard psychophysical procedure (for an introduction, see Watt, 1991). To perform this procedure of model fitting and parameter extraction for the data from Experiment 1 requires a model that relates the SOA to the probability of a touch-first response, and two candidate models suggest themselves.

The first model (the optimal inference model) assumes that (1) the participants have two categories of response, vision first or touch first, (2) there is uncertainty associated with the measurements of the time of onset of the tactile and visual stimuli, (3) this uncertainty is assumed to be approximately the same for both visual and tactile stimuli and is well described by a normal distributed internal estimate, and (4) participants perform hypothesis testing to determine which event occurred first: That is, they attempt to calculate the probability that the stimulus ensemble was of the type *touch first*. Given these four assumptions and an equally poor probability for it being either vision or touch first, probability theory (Bayes rule) states that $P(\text{touch first} \mid \text{SOA}) = P(\text{SOA} \mid \text{touch first})/[P(\text{SOA} \mid \text{touch first}) + P(\text{SOA} \mid \text{vision first})]$. In other words, to calculate which event occurred first, participants look at which of the two hypotheses (vision first or touch first) was more likely to have generated the observed SOA. Using the assumed normality of internal uncertainty, simple algebra gives us an equation $P(\text{touch first} \mid \text{SOA}) = 1/[1+\exp(-\text{slope} \cdot \text{SOA})]$—a logistic sigmoid model. A more complete de-
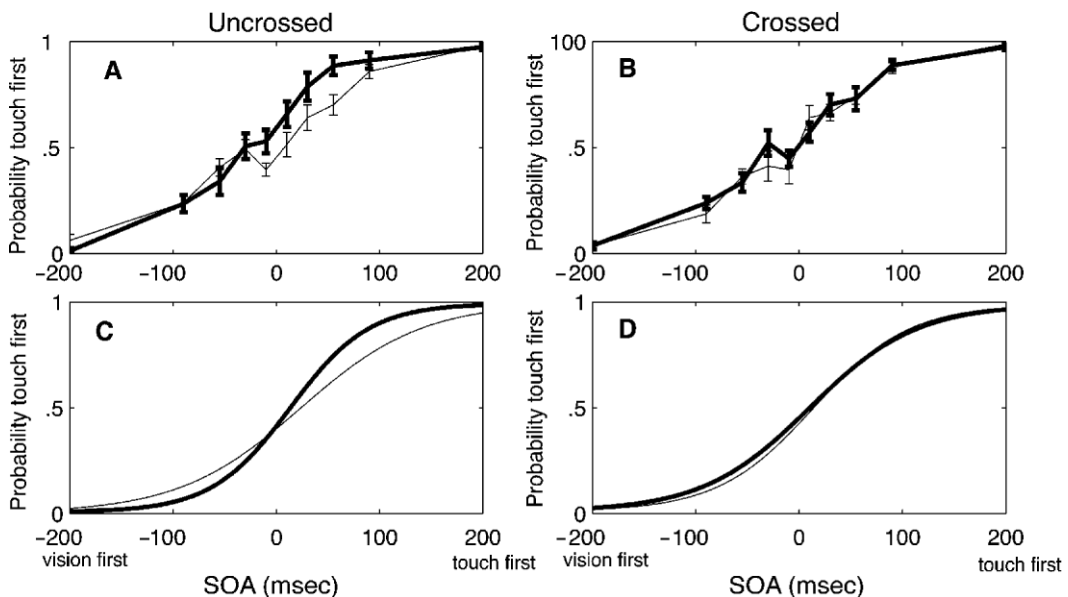
Figure 1. The population average response, the best fits, and the logic behind the fitting procedure for Experiment 1. (A) The probability of a *touch-first* response as a function of the stimulus onset asynchrony (SOA, in milliseconds; negative SOAs indicate that vision was presented first, whereas positive values indicate that touch was presented first) for participants adopting the uncrossed-hands posture. Thin lines represent responses made on the same-position trials, and thick lines represent responses made on the different-position trials (error bars correspond to the standard errors of the mean). (B) The same data, but this time for participants maintaining the crossed-hands posture. (C and D) The psychometric functions corresponding to panels A and B, with group average parameters (see the text for details of the fitting procedure adopted).
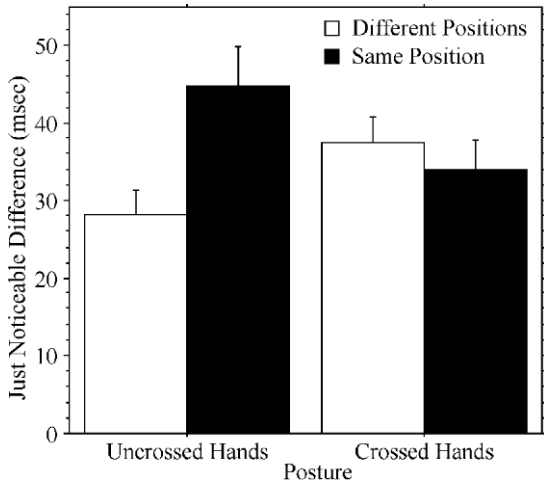
Figure 2. Accuracy of multisensory temporal order judgments (indicated by the just noticeable difference) when the participants made *which modality came first?* judgments for pairs of visual and tactile stimuli presented to either the same or different sides, with the hands placed in either an uncrossed or a crossed posture. Error bars represent the between-subjects standard errors of the mean.

rivation of this Bayes optimal decision rule under normally distributed noise can be found in Bishop (1995).

The alternative model simply assumes that the participants internally calculate the SOA and, if this is greater than some threshold, then they respond *touch first*. If participants' measurement of time was perfectly accurate, this would result in a step function relating the probability of responding to SOA, with the threshold at the PSS. If, more realistically, the estimated SOA is corrupted by noise, this would result in a noisy threshold model. Given that the internal noise is normally distributed, the resulting model is the well-known probit model (Finney, 1964). Although these two models have different motivations (optimal classification vs. noise corrupted threshold), they give essentially the same predictions. Both give S-shaped functions relating SOA to probability of touch first (the curves are within 5% of each other), and both predict that the slope of the curve is determined by the level of noise or uncertainty (high noise or uncertainty results in shallower sloped curves). For computa-

tional simplicity, therefore, we have chosen to use the optimal-inference–based logistic sigmoid model.

Two other considerations need to be taken into account when fitting the data. The first is that on a small fraction of trials (the blink rate), the participant can get distracted (or blink) and respond independently of the actual stimuli presented. If we assume this distracted responding is random (i.e., equally likely to be vision first or touch first), this percentage of distracted responses can be estimated by using a combined model in which, with a probability of (1 − blink rate), they respond as given in the above models and, with a probability of (blink rate), they respond randomly touch first or vision first. In Experiment 1, the proportion of such trials was estimated to be very small (1.3%), and so ignoring this effect would not affect the fit. However, in situations in which the number of trials in which the participant's blink rate is higher, ignoring such an effect can lead to a systematic overestimate of the slope parameter (cf. Swanson & Birch, 1992).

The last issue to note here is the optimization criteria used to find the best-fitting curve. The most natural criterion to adopt is to maximize the log-likelihood (i.e., the probability that the data were generated with a given set of parameters). The parameters that maximized the log-likelihood were estimated using a general-purpose nonlinear optimization algorithm, the Nelder–Mead simplex method (Nelder & Mead, 1965), implemented by the Matlab command "fmins" (code available on request).

The above arguments result in a model of the form

$$P(\text{response} | \text{SOA}) = \frac{(1 - \text{blink rate})}{1 + \exp[-slope * (\text{SOA} - \text{PSS})]} + \text{blink rate} * 0.5.$$

From the maximum likelihood fit of this model to participant's responses, the PSS, blink rate, and the JND (the interval between stimuli required for participants to judge which modality came first on 75% of the trials) was calculated (see Table 1). The latter is simply a function of the slope: $\text{JND}_{75} = 1.099/slope$.

A two-way within-subjects analysis of variance (ANOVA) performed on the PSS data derived from the psychometric curve-fitting had the factors of stimulus position (same

Table 1
Means and Standard Errors of the Means for Points of Subjective Simultaneity (PSS),
Slope, and Blink Rate Based on Fits to Each Individual Participant's Data,
as a Function of Posture and Stimulus Position in Experiment 1

	Posture							
	Uncrossed Hands				Crossed Hands			
	Same Position		Different Positions		Same Position		Different Positions	
Measure	M	SEM	M	SEM	M	SEM	M	SEM
PSS (msec)	−10.15	7.50	−23.73	10.01	−9.87	7.63	−13.95	5.65
Slope (msec)	0.017	0.002	0.026	0.003	0.022	0.002	0.019	0.002
Blink rate (lapses/trial)	0.006	0.004	0.014	0.011	0.022	0.011	0.021	0.019

Note—Negative values for the PSS indicate that the visual stimulus was presented first. The values were determined by averaging the PSS from the best-fitting function for each individual participant.

vs. different) and posture (uncrossed vs. crossed hands). This analysis revealed no significant effects [for stimulus position, $F(1,9) = 1.3, p = .29$; for posture, $F(1,9) = 3.5, p = .09$; and for the stimulus position $\times$ posture interaction, $F(1,9) < 1$, n.s.]. Visual inspection of Table 1 reveals that visual stimuli had to lead tactile stimuli (by approximately 14 msec, on average) for simultaneity to be reached. A t test revealed this asynchrony value to be significantly different from 0 [$t(9) = -2.5, p < .05$]. This finding, which has also been reported in several previous TOJ studies (e.g., Dinnerstein & Zlotogura, 1968; Spence, Shore, & Klein, 2001, Experiment 1), probably reflects the fact that conduction latencies for visual stimuli are longer than those for tactile stimuli and, so, visual stimuli had to be presented earlier than tactile stimuli in order to be perceived simultaneously.

A similar ANOVA performed on the JND data revealed a significant effect of stimulus position [$F(1,9) = 8.7, p = .02$] and an interaction between stimulus position and posture [$F(1,9) = 21.0, p = .001$; the main effect of posture was not significant, $F(1,9) < 1$, n.s.]. This interaction was decomposed using Bonferroni-corrected t tests [$t_{crit}(9) = 3.11$], comparing the two different stimulus positions for each posture and the two postures for each position. These revealed a significant effect of stimulus position for the uncrossed posture [$t(9) = 5.36, p < .0005$], but not for the crossed posture [$t(9) = 1.10$, n.s.]. Posture had a significant effect on stimuli from different positions [$t(9) = 8.19, p < .0001$], but not on stimuli from the same position [$t(9) = 2.98$, n.s.; see Figure 2].

Discussion

Although the results of Experiment 1 clearly demonstrate that spatial redundancy can facilitate visuotactile TOJs, we thought it possible that it might not necessarily play the same role for different pairs of stimulus modalities. In particular, one could argue that spatial redundancy might not play such an important role for audiovisual TOJs, given the well-known superior temporal resolution found in the auditory system over either the visual or the tactile system (e.g., Welch, DuttonHurt, & Warren, 1986). We also thought it important to replicate the beneficial effects of spatial redundancy on audiovisual TOJs, given that the majority of previous TOJ studies have used this pairing of stimulus modalities, rather than the visuotactile pairing used in Experiment 1. Therefore, in Experiment 2, we assessed the precision of *audiovisual* TOJs as a function of whether the auditory and the visual stimuli were presented from the same or different spatial positions. No postural manipulation was included in this new experiment, since the participants no longer received any tactile stimuli.

EXPERIMENT 2

Method

Participants. Ten participants (5 male and 5 female, mean age of 25 years) were recruited to take part in the experiment, which took approximately 55 min to complete. All the participants were

naive as to the purpose of the experiment, and none had taken part in the previous experiment.

Apparatus, Materials, Design, and Procedure. The apparatus, materials, design, and procedure were exactly as in Experiment 1, with the following exceptions. Two identical loudspeaker cones (VE100AO) were positioned so that one was directly behind each of the peripheral target LEDs. Auditory stimuli consisted of the presentation of an 8-msec white noise burst (82 dB[A] as measured from the participants' head position), and the visual stimuli consisted of the onset of either peripheral LED for 8 msec. We used somewhat shorter stimulus presentation durations in this experiment to ensure that the auditory and the visual stimuli were never presented simultaneously (as occurred for the shortest SOA in the previous experiment). White noise was presented continuously at 75 dB(A) throughout the experiment from a centrally placed loudspeaker cone to mask any noises made by the participants. The participants responded using a hand-held response pad. The same range of 10 SOAs between target stimuli was used in Experiment 2 (e.g., -200, -90, -55, -30, -10, $+10$, $+30$, $+55$, $+90$, and $+200$ msec), although positive SOAs now indicated that the auditory stimulus (rather than the tactile stimulus) was presented first, whereas negative values still indicated that the visual stimulus was presented first. The participants normally kept both thumbs on two separate keys placed vertically one above the other on a custom-made response pad (counterbalanced across participants). The participants were instructed to press the lower key to indicate that the auditory stimulus had been presented first, and the upper key whenever the visual stimulus appeared to have been presented first. Response feedback consisted of the illumination of one of the two LEDs placed directly above and below the fixation light immediately after a response was detected. The lower light indicated a *sound-first* response, whereas the upper light indicated a *vision-first* response.

Results

Best-fitting functions were used to calculate the PSS, slope, JND, blink rate, and fit for each participant for each condition (see Experiment 1 for details). The data from 2 of the participants were removed from subsequent data analysis because they had blink rates in one or more conditions that were far in excess of those of the other participants (i.e., mean proportion of blink trials of 37.3% and 55.7%, as compared with an average of 2.0% for the remainder of the participants). The results from the remaining 8 participants are shown in Figure 3 and Table 2.

Visual inspection of Table 2 reveals that visual stimuli had to lead auditory stimuli (by approximately 20 msec, on average) for simultaneity to be reached. This finding, which has also been reported in several previous TOJ studies (e.g., Gengel & Hirsh, 1970), probably reflects the fact that conduction latencies for visual stimuli are longer than those for auditory stimuli and, so, visual stimuli have to be presented first for them to be perceived as being simultaneous. A paired samples t test performed on the PSS data revealed no significant difference in the amount by which vision had to lead in the same- versus different-position conditions for the PSS to be achieved [$t(7) = -1.69, p = .14$]. However, the PSS was significantly different from 0 only for the different-position data [$t(7) = 2.95, p = .02$], and not for the same-position data [$t(7) = 1.5, p = .17$].

A one-tailed paired samples t test on the JND data revealed that the precision of multisensory TOJs was sig-
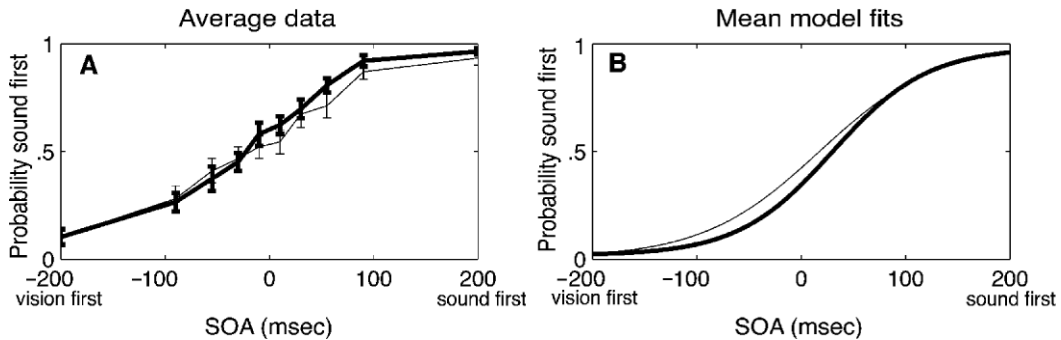
Figure 3. The population average response and the best fits for Experiment 2. (A) The probability of a *sound first* response as a function of the stimulus onset asynchrony (SOA, in milliseconds; negative SOAs indicate that vision was presented first, whereas positive values indicate that sound was presented first) and condition (same vs. different positions). Thin lines represent responses made on the same-position trials, and thick lines represent responses made on the different-position trials (error bars correspond to the standard errors of the mean). (B) The psychometric functions corresponding to panel A, with group average parameters.

nificantly better when the target stimuli were presented from different positions ($M = 53.3$ msec), rather than from the same position [$M = 41.5$ msec; $t(7) = 2.03$, $p < .05$; 7 out of the 8 participants exhibited this pattern of performance], just as in the uncrossed posture condition of the previous visuotactile experiment.

Discussion

The results of Experiment 2 replicated the main finding of our previous study—namely, that multisensory TOJs are more precise (i.e., JNDs are lower) when the individual stimuli are presented from different positions ($M = 41.5$ msec) rather than from the same position ($M = 53.3$ msec). This result further supports our main claim that previous studies may have systematically overestimated the precision with which people can make multisensory TOJs.

One reviewer raised the possibility that we might have underestimated the actual precision with which participants can make multisensory TOJs, because of the relatively wide range of SOAs used in the present experiments (e.g., compare our use of SOAs in the range of ±200 msec for our inexperienced psychophysical observers with Hirsh & Sherrick's, 1961, use of a range of only ±30 msec for their experienced observers). To assess the reviewer's point directly, we conducted a subsequent study in which 6 additional participants took part in a replication of Experiment 2, where the only change was a reduction in the range of SOAs used (−90, −75, −55, −30, −10, +10, +30, +55, +75, and +90 msec).

None of the participants in this new study was able to perform this task at much above chance level. Their performance was so bad, in fact, that it proved difficult to make meaningful fits of their data. Nevertheless, a simple comparison of the overall percentages of correct responses across the eight SOAs shared between the two experiments revealed particularly dramatic decrements in participants' performance when this narrower range

of SOAs at the 55- and 90-msec intervals was used. This result suggests that the use of a relatively wide range of SOAs in Experiments 1 and 2 was appropriate given the relative psychophysical inexperience of our participants. This result also suggests that the accuracy of multisensory TOJs, at least when assessed using the method of constant stimuli, can be dramatically influenced by the range of SOAs over which performance is examined (cf. Sade & Spitzer, 1998; Urbach & Spitzer, 1995). As such, any attempt to make direct comparison across different TOJ studies regarding the accuracy of multisensory temporal precision may be of only equivocal merit if those studies happen to have involved different ranges of SOAs.

GENERAL DISCUSSION

The results of both Experiment 1 (uncrossed posture) and Experiment 2 clearly demonstrate that introducing a spatially redundant cue can dramatically improve people's ability to determine the temporal order in which stimuli are presented to different sensory modalities. We

Table 2
Means and Standard Errors of the Means for Points of Subjective Simultaneity (PSS), Slope, Just Noticeable Difference (JND), and Blink Rate Based on Fits to Each Individual Participant's Data, as a Function of Stimulus Position in Experiment 2

Measure	Same Position		Different Positions	
	M	*SEM*	*M*	*SEM*
PSS (msec)	−10.3	6.7	−30.1	10.2
Slope (msec^{-1})	0.018	0.003	0.022	0.004
JND (msec)	53.3	15.4	41.5	9.8
Blink rate (lapses/trial)	0.005	0.004	0.034	0.012

Note—Negative values for the PSS indicate that the visual stimulus was presented first. The values were determined by averaging the PSS from the best-fitting function for each individual participant.

have demonstrated that spatial redundancy facilitates the precision with which people can make both visuotactile (Experiment 1) and audiovisual (Experiment 2) TOJs. These results are important because they highlight the spatial confound potentially inherent in nearly *all* previous multisensory TOJ studies, in which stimuli in different modalities were presented from different spatial locations (e.g., Hamlin, 1895; Hirsh & Sherrick, 1961; Rutschmann & Link, 1964; Smith, 1933; Teatini et al., 1976). Our results suggest that people are significantly worse at making multisensory TOJs than previous studies have suggested.

The present experiments may actually have underestimated the magnitude of the spatial redundancy effect, given that there was no unique location associated with either sensory modality (i.e., vision and touch could both be presented from either location on any trial, and the same was true for auditory and visual stimuli in Experiment 2). This contrasts with previous multisensory TOJ studies, in which a particular location has always been uniquely associated with stimuli in a particular sensory modality (i.e., auditory stimuli were always presented from headphones, and visual stimuli were always presented from the same position on the screen). Another reason our results may still reflect an overestimate of people's actual ability to make crossmodal TOJs is that other physical differences between the stimuli presented in each modality (such as their onset profile, their intensity, or their perceived duration) may also provide participants with redundant cues on which to base their response. We believe that it may, in principle, be impossible to ascertain a genuine threshold for multisensory TOJs, given the difficulty associated with ruling out *all* such sources of redundant cuing. Perhaps the most pragmatic solution is to try and reduce or eliminate as many of the possible sources of redundancy as possible. This raises the question of whether the inclusion of any other redundant dimension (such as differences in stimulus intensity or duration between the modalities) would also serve to facilitate multisensory TOJ performance or whether, instead, spatial location may be somehow special (perhaps resulting in emergent features, such as apparent motion, that would not occur following the inclusion of other sorts of redundant information; although see note 3). This issue remains as an important question for future research.

Multisensory Binding

If stimuli in different sensory modalities occur at the same spatial location at approximately the same time, it is likely that they refer to the same perceptual object. The brain uses spatiotemporal coincidence to determine which sensory stimuli should be bound together to form the multimodal perceptual objects of everyday life (cf. Driver & Spence, 2000; Neisser, 1976; Robertson et al., 2001). However, significant differences in the time required to transduce stimuli perceived by the various sensory epithelia mean that it is unlikely that information

from different modalities will actually arrive at the relevant cortical structures simultaneously (despite the fact that they may have occurred at the same time). For example, it takes less than 1 msec for auditory stimuli to be transduced at the cochlea, as compared with 30–40 msec for visual stimuli at the retina, with transduction latencies for tactile stimuli falling somewhere in between, depending on the body site stimulated (e.g., King & Palmer, 1985; Pöppel, Schill, & von Steinbüchel, 1990). Furthermore, given that sound and light travel through air at very different speeds, the asynchrony in arrival time of different sensory signals can be even more dramatic (consider the asynchrony that occurs when we perceive distant thunder and lightening; Pöppel, 1985/1988; Stone et al., 2001; although see Dennett, 1992). The relative slowness of transmission of sound through air is offset to some extent by the faster neural transduction of auditory stimuli once they have reached the sensory receptor surface. It has been claimed that these two factors cancel each other out at a distance of approximately 10 m, a distance known as the *horizon of simultaneity* (Pöppel et al., 1990). However, it is clear that for the majority of events not occurring at (or close to) the horizon of simultaneity, auditory and visual stimuli will arrive asynchronously.

Neurophysiologists studying the behavior of neurons in multisensory brain structures, such as the superior colliculus, have shown that cells involved in multisensory integration typically have fairly wide temporal windows for the integration of sensory signals to resolve this problem (see, e.g., King & Palmer, 1985; Meredith & Stein, 1983; Stein & Meredith, 1993). The problem of cortical synchronization is not restricted to multimodal perception but has also been reported for the identification of different visual features of an object (e.g., Anllo-Vento & Hillyard, 1996; Moutoussis & Zeki, 1997; Nishida & Johnston, 2002; see also Schröger & Wolff, 1997, and Widmann, Scharf, & Schröger, 2000, for related work on auditory feature processing). It has been argued that the same mechanisms/processes postulated to resynchronize asynchronous sensory signals occurring within a particular sensory modality may also be used to resynchronize stimuli presented to different sensory modalities (Grossberg & Grunewald, 1997).

One consequence of binding different sensory signals into multisensory perceptual objects or events may be that information regarding the relative temporal onsets of the component stimuli are lost. In other words, it appears that when multisensory binding takes place, the various sensory signals are "ventriloquized in time" to provide a single temporal onset for the multisensory percept (cf. Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2001; Scheier, Nijhawan, & Shimojo, 1999). This form of temporal ventriloquism may be equivalent to the well-known *spatial* ventriloquism effect, in which spatially discrepant stimuli in different modalities are bound together so that they appear to originate from the same spatial location, at least when

the spatial discrepancy is not too great (e.g., Bertelson, 1998; Bertelson & Aschersleben, 2001; Slutsky & Recanzone, 2001; Welch & Warren, 1980). Similarly, Piéron (1952) and Fraisse (1963) have also reported that pairs of stimuli presented within the same sensory modality are more likely to be perceived as simultaneous when they are presented from the same, rather than from different, spatial locations.

The superior performance observed when the multisensory stimuli come from different locations stands in marked contrast to previous *unimodal* studies, which have typically shown better performance for discriminating multiple visual attributes (such as length and shape) if they refer to the same static object (e.g., Duncan, 1984; Egly, Driver, & Rafal, 1994; O'Craven, Downing, & Kanwisher, 1999). The standard account of this one-object benefit for unimodal visual attributes involves object-based mechanisms of attention. The present results suggest that these benefits may come at the cost of a loss of temporal order resolution regarding the relative temporal onsets of those attributes. Testing this claim in relation to the perception of different visual attributes of an object will be an interesting task for future research.

The Cost of Crossing the Hands

The results of the crossed-hands posture condition of Experiment 1 demonstrate that participants are unable to take advantage of the spatial redundancy gain (i.e., performance on different-position trials was no better than that on same-position trials) when they cross their hands over the midline. When participants adopted the crossed-hands posture, performance was intermediate between the performances reported for the same- and the different-position conditions in the uncrossed posture condition. Given that previous research has shown that people's ability to order/discriminate visual stimuli is unaffected by crossing the hands (e.g., Shore et al., 2002a; Yamamoto & Kitazawa, 2001a), our results suggest that participants may instead be confused about the location of the tactile stimulus on some percentage of the crossed-hands trials (see Shore et al., 2002a; Yamamoto & Kitazawa, 2001a), the reasoning being that if participants are unsure of which hand received the tactile stimulus when they cross their hands, they may confound same- and different-location trials, resulting in an intermediate level of temporal discrimination performance. This difference between the senses may be attributable to differences in the coordinate framework in which spatial coding takes place for visual and tactile stimuli: Whereas extrinsic spatial (i.e., allocentric) coordinates are typically used by the visual system, the haptic system frequently processes spatial information in body-centered coordinates instead (e.g., Millar, 1994; O'Connor & Hermelin, 1978; Röder, Spence, & Rösler, 2002).

Although this account must, at present, remain speculative, it nevertheless seems likely that confusions regarding the location of tactile stimuli with the crossed-

hands posture are restricted to situations in which two stimuli are presented in quick succession, given that previous research has shown that people have little problem determining the veridical location of single tactile stimuli presented to one or other hand when the hands are crossed (e.g., Anzola et al., 1977; Bradshaw et al., 1992; see Shore et al., 2002a, for a more detailed discussion of this point). If correct, this would argue that people may be confused about the location of tactile stimuli even when such tactile stimuli are presented together with a near-simultaneous visual event by the other hand.

By combining the TOJ methodology outlined here with cognitive neuroimaging techniques—for example, positron emission tomography, functional magnetic resonance imaging, and such electrophysiological measures as event-related potentials—future studies will hopefully provide a better understanding of the network of neural structures, such as the basal ganglia, cerebellum, and insula, underlying temporal perception and temporal discrimination performance (e.g., Bushara, Grafman, & Hallett, 2001; Raizada & Poldrack, 2001; Rao, Mayer, & Harrington, 2001). It is particularly important that the appropriate methods be developed for indexing multisensory temporal perception, given the interest shown by clinical researchers in the potential application of multisensory temporal perception measures as predictors of poor reading ability (e.g., Dinnerstein & Blitz, 1964; Rose, Feldman, Jankowski, & Futterweit, 1999; see also Carmon & Nachshon, 1971).

REFERENCES

ALLEN, P. G., & KOLERS, P. A. (1981). Sensory specificity of apparent motion. *Journal of Experimental Psychology: Human Perception & Performance,* **7**, 1318-1326.

ANLLO-VENTO, L., & HILLYARD, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics,* **58**, 191-206.

ANZOLA, G. P., BERTOLONI, G., BUCHTEL, H. A., & RIZZOLATTI, G. (1977). Spatial compatibility and anatomical factors in simple and choice reaction time. *Neuropsychologia,* **15**, 295-302.

BALD, L., BERRIEN, F. K., PRICE, J. B., & SPRAGUE, R. O. (1942). Errors in perceiving the temporal order of auditory and visual stimuli. *Journal of Applied Psychology,* **26**, 382-388.

BERTELSON, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science, Vol. 2. Biological and cognitive aspects* (pp. 419-439). Hove, U.K.: Psychological Press.

BERTELSON, P., & ASCHERSLEBEN, G. (2001, January). *Crossmodal interaction on the time dimension: Evidence from auditory–visual time order judgment.* Paper presented at the London meeting of the Experimental Psychology Society.

BISHOP, C. M. (1995). *Neural networks for pattern recognition.* Oxford: Oxford University Press, Clarendon Press.

BLOCH, A. M. (1887). Exnerschen ähnliche Versuche mit analogen Ergebnissen [Replication of Exner's studies with analogous results]. *Revue Scientifique,* **XXXIX**, S.585.

BRADSHAW, J. L., HOWARD, M. J., PIERSON, J. M., PHILLIPS, J., & BRADSHAW, J. A. (1992). Effects of expectancy and attention in vibrotactile choice reaction time tasks. *Quarterly Journal of Experimental Psychology,* **44A**, 509-528.

BURTT, H. E. (1917). Tactile illusions of movement. *Journal of Experimental Psychology,* **2**, 371-385.

BUSHARA, K. O., GRAFMAN, J., & HALLETT, M. (2001). Neural corre-

lates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience*, **21**, 300-304.

CARMON, A., & NACHSHON, I. (1971). Effect of unilateral brain damage on perception of temporal order. *Cortex*, **7**, 410-418.

COREN, S., WARD, L. M., & ENNS, J. T. (1999). *Sensation and perception* (5th Ed.). Fort Worth: Harcourt Brace.

DENNETT, D. C. (1992). Temporal anomalies of consciousness. In Y. Christen & P. S. Churchland (Eds.), *Neurophilosophy and Alzheimer's disease*. Berlin: Springer-Verlag.

DINNERSTEIN, A. J., & BLITZ, B. (1964). Perceptual delays and speed of reading and tapping. *Perceptual & Motor Skills*, **19**, 867-873.

DINNERSTEIN, A. J., & ZLOTOGURA, P. (1968). Intermodal perception of temporal order and motor skills: Effects of age. *Perceptual & Motor Skills*, **26**, 987-1000.

DRIVER, J., & SPENCE, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, **10**, R731-R735.

DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General,* **113,** 501-517.

EFRON, R. (1963). The effect of handedness on the perception of simultaneity and temporal order. *Brain*, **86**, 261-284.

EGLY, R., DRIVER, J., & RAFAL, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, **123**, 161-177.

ESTES, W. K. (1956). The problem of interference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.

EXNER, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Processe [Experimental study of the most simple psychological processes]. *Archiv für die geschichte Physiologie (Pflüger's Archive)*, **11**, 403-432.

FENDRICH, R., & CORBALLIS, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, **63**, 719-725.

FINNEY, D. J. (1964). *Probit analysis: Statistical treatment of the sigmoid response curve* (2nd ed.). Cambridge: Cambridge University Press.

FRAISSE, P. (1963). *The psychology of time*. Evanston, NY: Harper & Row.

GENGEL, R. W., & HIRSH, I. J. (1970). Temporal order: The effect of single versus repeated presentations, practice, and verbal feedback. *Perception & Psychophysics*, **7**, 209-211.

GROSSBERG, S., & GRUNEWALD, A. (1997). Cortical synchronization and perceptual framing. *Journal of Cognitive Neuroscience*, **9**, 117-132.

GUINZBERG, R. L. (1928). È possibile l'apprendimento di sensazioni eterogenee come perfettamente simultanee? (Is it possible to learn that heterogenous sensations are perfectly simultaneous?) *Archivi Italiani di Psicologia*, **6**, 103-114.

HAMLIN, A. J. (1895). On the least observable interval between stimuli addressed to disparate senses and to different organs of the same sense. *American Journal of Psychology*, **6**, 564-575.

HIRSH, I. J. (1988). Auditory perception and speech. In R. C. Atkinson, R. J. Herrnstein, G. Lindsey, & R. D. Luce (Eds.), *Stevens' Handbook of experimental psychology: Perception and motivation* (2nd ed., pp. 377-408). New York: Wiley.

HIRSH, I. J., & FRAISSE, P. (1964). Simultanéité et succession de stimuli hétérogènes [Simultaneity and succession of heterogeneous stimuli]. *L'Année Psychologique*, **64**, 1-19.

HIRSH, I. J., & SHERRICK, C. E, JR.. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, **62**, 423-432.

JAŚKOWSKI, P., JAROSZYK, F., & HOJAN-JERIERSKA, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, **52**, 35-38.

KENNETT, S., EIMER, M., SPENCE, C., & DRIVER, J. (2001). Tactile–visual links in exogenous spatial attention under different postures: Convergent evidence from psychophysics and ERPs. *Journal of Cognitive Neuroscience*, **13**, 462-478.

KENNETT, S., SPENCE, C., & DRIVER, J. (2002). Visuo-tactile links in covert exogenous spatial attention remap across changes in unseen hand posture. *Perception & Psychophysics*, **64**, 1083-1094.

KING, A. J., & PALMER, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, **60**, 492-500.

McFARLAND. D. J., CACACE, A. T., & SETZEN, G. (1998). Temporal-order discrimination for selected auditory and visual stimulus dimensions. *Journal of Speech, Language, & Hearing Research*, **41**, 300-314.

MEREDITH, M. A., & STEIN, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, **221**, 389-391.

MILLAR, S. (1994). *Understanding and representing space: Theory and evidence from studies with blind and sighted children*. Oxford: Oxford University Press.

MOREIN-ZAMIR, S., SOTO-FARACO, S., & KINGSTONE, A. F. (2001). Captured vision: Sounds influence the perception of visual temporal order. *Abstracts of the Psychonomic Society*, **6**, 91.

MOUTOUSSIS, K., & ZEKI, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London: Series B*, **264**, 393-399.

NEISSER, U. (1976). *Cognition and reality*. San Francisco: Freeman.

NELDER, J. A., & MEAD, R. (1965). A simplex method for function minimisation. *Computer Journal*, **7**, 308-313.

NISHIDA, S., & JOHNSTON, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology*, **12**, 359-368.

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

O'CRAVEN, K. M., DOWNING, P. E., & KANWISHER, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, **401**, 584-587.

PIÉRON, H. (1952). *The sensations: Their functions, processes and mechanisms*. New Haven: Yale University Press.

PÖPPEL, E. (1985). *Grenzen des Bewußtseins*. Stuttgart: Deutsche Verlags-Anstal. [Translated as *Mindworks: Time and conscious experience* (1988). New York: Harcourt Brace Jovanovich.]

PÖPPEL, E., SCHILL, K., & STEINBÜCHEL, N. VON (1990). Sensory integration within temporally neutral system states: A hypothesis. *Naturwissenschaften*, **77**, 89-91.

RAIZADA, R. D. S., & POLDRACK, R. A. (2001). Event-related fMRI of audio-visual simultaneity perception. *Society for Neuroscience Abstracts*, Poster No. 511.14.

RAO, S. M., MAYER, A. R., & HARRINGTON, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, **4**, 317-323.

ROBERTSON, G. E., HAIRSTON, W. D., WALLACE, M. T., STEIN, B. E., LAURIENTI, P. J., & SCHIRILLO, J. A. (2001). Unifying multisensory signals across time and space. *Society for Neuroscience Abstracts*, Poster No. 511.19.

RÖDER, B., SPENCE, C., & RÖSLER, F. (2002). Assessing the effect of posture changes on tactile inhibition of return. *Experimental Brain Research*, **143**, 453-462.

ROSE, S. A., FELDMAN, J. F., JANKOWSKI, J. J., & FUTTERWEIT, L. R. (1999). Visual and auditory temporal processing, cross-modal transfer, and reading. *Journal of Learning Disabilities*, **32**, 256-266.

RUTSCHMANN, J., & LINK, R. (1964). Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual & Motor Skills*, **18**, 345-352.

SADE, A., & SPITZER, H. (1998). The effects of attentional spread and attentional effort on orientation discrimination. *Spatial Vision*, **11**, 367-383.

SCHEIER, C. R., NIJHAWAN, R., & SHIMOJO, S. (1999). Sound alters visual temporal resolution [Abstract]. *Investigative Ophthalmology & Visual Science*, **40**, S792.

SCHRÖGER, E., & WOLFF, C. (1997). Fast preattentive processing of location: A functional basis for selective listening in humans. *Neuroscience Letters*, **232**, 5-8.

SHERRICK, C. E. (1976). The antagonisms of hearing and touch. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, & S. R. Silverman (Eds.), *Hearing and Davis: Essays honoring Hallowell Davis* (pp. 149-158). St. Louis: Washington University Press.

SHORE, D. I., SPENCE, C., & KLEIN, R. M. (2001). Visual prior entry. *Psychological Science*, **12**, 205-212.

SHORE, D. I., SPRY, E., & SPENCE, C. (2002a). Confusing the mind by crossing the hands. *Cognitive Brain Research*, **14**, 153-163.

SHORE, D. I., SPRY, E., & SPENCE, C. (2002b). *Spatial modulation of tactile temporal order judgments*. Manuscript submitted for publication.

SIMON, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus–response compatibility* (pp. 31-86). Amsterdam: Elsevier.

SLUTSKY, D. A., & RECANZONE, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, **12**, 7-10.

SMITH, W. F. (1933). The relative quickness of visual and auditory perception. *Journal of Experimental Psychology*, **16**, 239-257.

SPENCE, C., & DRIVER, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, **59**, 389-403.

SPENCE, C., NICHOLLS, M. E. R., & DRIVER, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, **63**, 330-336.

SPENCE, C., NICHOLLS, M. E. R., GILLESPIE, N., & DRIVER, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, **60**, 544-557.

SPENCE, C., PAVANI, F., & DRIVER, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1298-1319.

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

STEIN, B. E., & MEREDITH, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

STONE, J. V., HUNKIN, N. M., PORRILL, J., WOOD, R., KEELER, V., BEANLAND, M., PORT, M., & PORTER, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London: Series B*, **268**, 31-38.

STRYBEL, T. Z., MANLIGAS, C. L., CHAN, O., & PERROTT, D. R. (1990). A comparison of the effects of spatial separation on apparent motion in the auditory and visual modalities. *Perception & Psychophysics*, **47**, 439-448.

SWANSON, W. H., & BIRCH, E. E. (1992). Extracting thresholds from noisy psychophysical data. *Perception & Psychophysics*, **51**, 409-422.

TAYLOR, B. (1978). *Dimensional redundancy in the processing of vibrotactile temporal order*. Unpublished doctoral dissertation, Princeton University.

TEATINI, G., FARNÈ, M., VERZELLA, F., & BERRUECOS, P., JR. (1976). Perception of temporal order: Visual and auditory stimuli. *Giornale Italiano di Psicologia*, **3**, 157-164.

URBACH, D., & SPITZER, H. (1995). Attentional effort modulated by task difficulty. *Vision Research*, **35**, 2169-2177.

VAN DE PAR, S., KOHLRAUSCH, A., & JUOLA, J. F. (1999). Judged synchrony/asynchrony for light–tone pairs. *Abstracts of the Psychonomic Society*, **4**, 60.

WATT, R. J. (1991). *Understanding vision*. London: Academic Press.

WEILAND, B. A. (1960). The interaction of space and time in cutaneous perception. *American Journal of Psychology*, **73**, 248-255.

WELCH, R. B., DUTTONHURT, L. D., & WARREN, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, **39**, 294-300.

WELCH, R. B., & WARREN, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **3**, 638-667.

WHIPPLE, G. M., SANFORD, E. C., & COLGROVE, F. W. (1899). Minor studies from the psychological laboratory of Clark University: On nearly simultaneous clicks and flashes. The time required for recognition: Notes on mental standards of length. *American Journal of Psychology*, **10**, 280-295.

WIDMANN, A., SCHARF, B., & SCHRÖGER, E. (2000). Frequency and location discrimination: Which is faster? In A. Schick, M. Meis, & C. Reckhardt (Eds.), *Contributions to Psychological Acoustics: Results of the Eighth Oldenburg Symposium on Psychological Acoustics* (pp. 373-380). Oldenburg: Bibliotheks- und Informationssystem der Universität Oldenburg.

YAMAMOTO, S., & KITAZAWA, S. (2001a). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience*, **4**, 759-765.

YAMAMOTO, S., & KITAZAWA, S. (2001b). Sensation at the tips of invisible tools. *Nature Neuroscience*, **4**, 979-980.

ZAMPINI, M., GUEST, S., SHORE, D. I., & SPENCE, C. (2003). *Audiovisual synchrony judgments*. Manuscript submitted for publication.

NOTES

1. Some researchers have also used a simultaneous/successive judgment task to investigate multisensory temporal processing (e.g., Efron, 1963; Exner, 1875; Guinzberg, 1928; Stone et al., 2001; van de Par, Kohlrausch, & Juola, 1999). Using this form of discrimination task, researchers can determine the point of subjective simultaneity of pairs of stimuli presented to different modalities by varying the stimulus onset asynchrony (SOA) between them. However, it has been argued that the simultaneous/successive task may not provide a reliable estimate of a participant's sensitivity to asynchrony (i.e., equivalent to the just noticeable difference in TOJ studies), given the likelihood that the criterion for simultaneity adopted by participants will be affected by the particular range of SOAs presented (Spence, Shore, & Klein, 2001; Zampini, Guest, Shore, & Spence, 2003).

2. Many different factors have been shown to dramatically affect the JND. These include the use of untrained psychophysical observers, as opposed to Hirsh and Sherrick's (1961) reliance on highly trained psychophysical observers (see Bald et al., 1942; Dinnerstein & Zlotogura, 1968; Gengel & Hirsh, 1970; Hirsh, 1988; Hirsh & Fraisse, 1964; Spence, Shore, & Klein, 2001), how often the stimulus pair is presented (note that the participants in Hirsh & Sherrick's study were allowed to listen repeatedly to the stimulus pair until certain of their TOJ response), the introduction of background white noise (Sherrick, 1976), the removal of corrective error feedback (Hirsh, 1988), and the task and context in which the stimuli are presented (e.g., Gengel & Hirsh, 1970; McFarland, Cacace, & Setzen, 1998).

3. The introduction of a spatial separation between stimuli in TOJ experiments may also facilitate performance, owing to the emergence of apparent motion cues that occur only when stimuli are rapidly presented from *different* spatial locations. Although there is extensive evidence for the occurrence of apparent motion within vision, audition, and touch (e.g., Burtt, 1917; Strybel, Manligas, Chan, & Perrott, 1990), researchers have argued that it does not occur when stimuli are presented in different sensory modalities (e.g., Allen & Kolers, 1981; Hirsh & Sherrick, 1961; McFarland et al., 1998).

4. There is, at present, no agreement in the literature as to what constitutes the most appropriate criterion for matching stimuli crossmodally—that is, should they be matched for simple detection latencies, for discrimination latencies, for subjective intensities, or according to some other criterion (for a sense of the variety of different solutions this problem has attracted, see, e.g., Bald et al., 1942; Hirsh & Sherrick, 1961; Spence & Driver, 1997; Spence, Shore, & Klein, 2001; Stone et al., 2001; Strybel et al., 1990; Whipple et al., 1899). However, it should be noted that any intensity-based modulation of the PSS present in our study is *orthogonal* to the experimental manipulations of interest (i.e., same- vs. different-position cost and uncrossed vs. crossed posture).

5. It is possible that the onset of the unimodal visual fixation light at the start of each trial may have acted as an exogenous cue (or priming stimulus) that automatically drew the participant's attention toward the visual modality (see Spence & Driver, 1997; and Spence, Nicholls, & Driver, 2001, on this issue). Although the 750-msec interval between the onsets of the fixation light and the first target stimulus is outside the range of many exogenous cuing effects, any residual effect should only have reduced the overall amount by which vision had to lead touch for the PSS to be achieved. Note also that the use of unimodal visual response feedback lights in this and many other studies may also have led to a residual attentional bias toward the visual modality (see Spence, Shore, & Klein, 2001).