Audio–visual simultaneity judgments

MASSIMILIANO ZAMPINI
University of Oxford, Oxford, England
and University of Trento, Rovereto, Italy

STEVE GUEST
University of Oxford, Oxford, England

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

and

CHARLES SPENCE
University of Oxford, Oxford, England

The relative spatiotemporal correspondence between sensory events affects multisensory integration across a variety of species; integration is maximal when stimuli in different sensory modalities are presented from approximately the same position at about the same time. In the present study, we investigated the influence of spatial and temporal factors on audio–visual simultaneity perception in humans. Participants made unspeeded simultaneous versus successive discrimination responses to pairs of auditory and visual stimuli presented at varying stimulus onset asynchronies from either the same or different spatial positions using either the method of constant stimuli (Experiments 1 and 2) or psychophysical staircases (Experiment 3). The participants in all three experiments were more likely to report the stimuli as being simultaneous when they originated from the same spatial position than when they came from different positions, demonstrating that the apparent perception of multisensory simultaneity is dependent on the relative spatial position from which stimuli are presented.

Historically, some researchers have argued that people do not experience a strong perception of simultaneity for stimuli presented to different sensory modalities (such as audition and vision; see, e.g., Fraisse, 1964; Guinzberg, 1928; Piéron, 1952). For example, Piéron (1952, p. 295) claimed that whereas a "rigorous impression" of simultaneity can occur for stimuli presented within a sensory modality, no clear sensation of simultaneity is perceived for stimuli presented to different sensory modalities. A similar point was made by Fraisse (1964, p. 109) when he stated that "it is very difficult to assess the simultaneity of two sensations which have nothing in common; this is true for stimulations involving the same sense, but even more so for heterogeneous stimulations."

On the other hand, contemporary research provides a number of demonstrations that appear to show that people can judge the apparent simultaneity of events across the senses (see Spence, Shore, & Klein, 2001, for a review). In fact, a growing body of behavioral and neurophysiological research now highlights the crucial role that temporal synchrony and spatial coincidence play in modulating the effects of multisensory integration (e.g., Lewald, Ehrenstein, & Guski, 2001; Slutsky & Recanzone, 2001; see Driver & Spence, 2000, and Stein & Meredith, 1993, for reviews). Research also suggests, however, that temporal and spatial factors may play somewhat distinct roles in modulating multisensory integration effects (for reviews, see Calvert, Brammer, & Iversen, 1998; Munhall & Vatikiotis-Bateson, 2004; Spence & Driver, 2004; Stein & Meredith, 1993). This is clearly true for the case of audiovisual speech perception, in which temporal desynchronization of the auditory and visual components has been shown to have a far more detrimental effect on speech perceptibility (e.g., McGrath & Summerfield, 1985; Pandev, Kunov, & Abel, 1986) than does spatial misaligment (e.g., Bertelson, Vroomen, & de Gelder, 1997; Jones & Munhall; 1997; Radeau & Bertelson, 1977).

Researchers have developed several experimental paradigms to assess the perception of multisensory simultaneity in humans, including the temporal order judgment (TOJ) task and the simultaneity judgment task. In a typical simultaneity judgment experiment, pairs of auditory and visual stimuli are presented at a range of differ-

ent stimulus onset asynchronies (SOAs), using the method of constant stimuli (e.g., Spence, Shore, & Klein, 2001), and participants are required to judge whether the stimuli were presented simultaneously or successively (e.g., Engel & Dougherty, 1971; Exner, 1875; Hirsh & Fraisse, 1964; Slutsky & Recanzone, 2001; Stone et al., 2001). Psychophysical analysis of the results of such studies is used to determine the SOA at which the participants would have been most likely to have made a "simultaneous" response, known as the *point of subjective simultaneity* (PSS).[1] In one recent study, Stone et al. (2001) reported that for the PSS to be reached, a light placed 50 cm in front of participants had to be illuminated an average of 50 msec before a sound was presented over headphones. Stone et al. also found that there was a great deal of between-participants variability in the PSS; for 1 participant the sound had to lead the light by 21 msec for simultaneity to be achieved, whereas for another the light had to lead the sound by 150 msec before the PSS was attained (cf. Guinzberg, 1928). These differences in the PSS between different participants were robust across testing sessions, leading Stone et al. to claim that what constitutes simultaneity for pairs of auditory and visual stimuli may be quite different for different people. This conclusion should come as little surprise to those familiar with the early literature in experimental psychology on the personal equation (e.g., Mollon & Perkins, 1996; Spence, Shore, & Klein, 2001).

An extensive body of research now shows that auditory and visual stimuli are far more likely to be bound into perceptual objects/events when presented from the same spatial location than from different locations (see, e.g., Bertelson & de Gelder, 2004, for a review). Consequently, one might expect that people would be more likely to judge two stimuli as simultaneous if they were presented from the same position rather than from different positions. However, we are aware of no previous studies in which direct attempts have been made to assess the effect of relative spatial position (i.e., same vs. different) on temporal perception, as assessed by the simultaneity task. Instead, the auditory and visual stimuli used in the majority of previous simultaneity studies have been presented only from fixed but differing locations (see, e.g., Hirsh & Fraisse, 1964; Stone et al., 2001, Experiment 1; Van de Par, Juola, & Kohlrausch, 1999), auditory stimuli typically being presented over headphones, whereas visual stimuli were presented in front of the participant (but see Engel & Dougherty, 1971, Slutsky & Recanzone, 2001, and Stone et al., 2001, Experiment 2, for exceptions).[2]

The aim of the present study was to examine whether the relative positions of auditory and visual stimuli modulate the perception of audio–visual simultaneity. To this end, pairs of auditory and visual stimuli were presented from either the same spatial location or different locations to either side of fixation, at varying SOAs, using the method of constant stimuli. On each trial, the participants

were required to make an unspeeded "simultaneous"–"successive" discrimination response regarding the pair of auditory and visual stimuli (no matter whether they were presented from the same spatial position or from different positions). We predicted that the participants would be more likely to respond "simultaneous" when the auditory and visual stimulus pairs were presented from the same position rather than from different positions. This prediction was based on the fact that there should be a greater tendency to bind stimuli from different sensory modalities together when they come from the same location rather than from different locations (see, e.g., Spence & Driver, 2004; Stein & Meredith, 1993). We predicted that this merging might blur the relative onsets of the individual sensory features, thereby increasing the likelihood of the perception of simultaneity rather than the perception of a sequence of two distinct unimodal events.

EXPERIMENT 1

Method

Participants. Twenty participants (mean age 26 years) took part in the experiment as paid volunteers. All of the participants were right-handed by self-report. Visual acuity was normal or corrected to normal, and all of the participants reported normal hearing. The participants were naive as to the purpose of the experiment and varied in their previous experience with psychophysical testing procedures. The experiment took approximately 50 min to complete. The experiment was conducted in accordance with the ethical standards of the American Psychological Association.

Apparatus and Stimuli. The experiment was conducted in a completely dark sound-attenuated room. A red light-emitting diode (LED) was placed on the table 62 cm in front of the participant and served as a fixation point. Similar LEDs were placed above and below this fixation point to indicate to the participant which response they had made. Two identical small loudspeaker cones (8 cm in diameter) were positioned 26 cm (24º) to either side of the fixation light, at the same distance from the participant. A red LED was placed directly in front of each of these eccentric loudspeaker cones. The auditory stimuli consisted of the presentation of a 9-msec burst of white noise [82 dB(A) as measured from the participant's head position], and the visual stimuli consisted of the onset of either peripheral LED for 9 msec. No specific attempt was made to match the intensities of the stimuli, which were clearly suprathreshold. Throughout the experiment, white noise was presented continuously at 75 dB(A) (as measured from the participant's head position) from two additional loudspeakers placed on the table in front of the participant to mask any noises made by the participant.

The participant normally kept both thumbs on two separate keys placed one above the other on a hand-held response pad (with one key held slightly closer to the participant). The participant was instructed to press the upper key to indicate that the two stimuli appeared to have been presented simultaneously, and the lower key whenever the stimuli appeared to have been presented asynchronously (or successively). The spatially compatible illumination of one of the two LEDs placed directly above and below the fixation light immediately after a response indicated to the participant which response he or she had made.

Design. There were two within-participants factors: stimulus position (same vs. different) and SOA (−200, −70, −30, −20, 0, 20, 30, 70, and 200 msec; negative SOAs indicate that the auditory

stimulus was presented first, whereas positive values indicate that the visual stimulus was presented first). The selection of this particular range of SOAs was based on our previous research on audiovisual TOJs (see Spence, Baddeley, Zampini, James, & Shore, 2003). An equal number of trials was presented at each SOA, with the exception that twice as many trials were presented at the 0-msec SOA as at any other SOA, in an attempt to ensure a sufficient number of apparently simultaneous stimulus pairs (although, given Stone et al.'s, 2001, findings, it should be noted that the 0-msec SOA may not actually have corresponded to the true point of phenomenal simultaneity—i.e., to the PSS). The auditory and visual stimuli were presented from each side with equal probability during each experimental session. The same-side and different-sides trials were also presented with equal probability. All the participants completed two blocks of 30 practice trials, followed by eight blocks of 80 experimental trials. The SOAs were doubled in the first practice block to facilitate the acquisition of the task by the participants.

Procedure. The fixation light was illuminated at the beginning of each trial. The participants were instructed to maintain their fixation on this central red LED throughout each block of trials. The first stimulus (or both stimuli on trials on which the SOA was 0 msec) was presented from the left and/or the right after a delay of 750 msec. The second stimulus was presented after the SOA specified for that particular condition (on trials on which the SOA was greater than 0 msec). The task was unspeeded, and the participants were informed that they should respond only when confident of their response (although within the 3,500 msec allowed before termination of the trial). If the participants responded prior to the onset of the first stimulus or failed to make a response before the trial was terminated (3,500 msec after the onset of the first stimulus), error feedback was presented. This consisted of the flickering of the fixation light for 1,000 msec. Such responses occurred on fewer than 1% of trials overall and were not analyzed. These trials were not repeated later in the experimental session. Otherwise, the participant's response was indicated by the illumination of one of the central feedback lights for 500 msec. The fixation light was illuminated to indicate the start of the next trial 750 msec after the end of the trial preceding it.

Results

The observed distribution of responses was fitted to a Gaussian function for each participant, using maximum likelihood estimation (see Myung, 2003, for a tutorial review). We compared the observed distribution with a normal distribution for each participant using the Kolmogorov–Smirnov goodness-of-fit test. All observations could reasonably have come from the specified distribution ($p > .05$),[3] and so all of the participants were considered in the subsequent analyses. The parameters obtained from each participant's data were averaged, and the curves described by these mean parameters are plotted in Figure 1, along with the empirical data points, again averaged across all participants. The participants were more likely to judge the two stimuli as being simultaneous if they were presented from the same position rather than from different positions across the entire range of SOAs tested.

The data were formally analyzed by comparison of the three defining parameters of the fitted distributions [i.e., the mean of the distribution, or PSS, the peak height of the distribution, and the standard deviation (SD) of the distribution] for the same-location data versus those obtained for the different-locations data, for each participant individually. These data are presented for each par-

ticipant in Table 1. Visual stimuli had to be presented further in advance of auditory stimuli for the PSS to be achieved when the two stimuli were presented from different spatial locations, in comparison with when they were presented from the same location ($M = 32.1$ and 19.4 msec for different spatial locations and same location, respectively). A t test comparing the PSS across the relative stimulus positions revealed that this difference was statistically significant [$t(19) = 4.2, p < .001$].

The SDs of the derived functions provide an estimate of the spread of the distribution and so are inversely related to the slope of the underlying psychometric function for each set of data. The SDs therefore provide one measure of how difficult participants found the simultaneity judgment task across the range of SOAs tested. Specifically, smaller SDs indicate steeper psychometric functions and thus better discriminative performance. The SD is monotonically related to the conventionally defined just noticeable difference; twice the SD represents a more conservative (84%) threshold than the traditional 75% value. Visual inspection of the data presented in Table 1 reveals that the SD is larger for the same-position data ($M = 137$) than for the different-positions data ($M = 117$) in 17 of the 20 participants. These values are presented in Table 1, along with the range of SOAs over which each participant had a greater than 50% likelihood of making a "simultaneous" response.

We formally compared the SDs of the fitted distributions between the same-position and different-positions conditions using the same method as for the PSS data. In this case, a t test revealed that the SDs were significantly larger overall for the same-position data than for the different-positions data [$t(19) = 3.7, p = .001$]. It is interesting to note that the individual PSS values appear somewhat less variable in our experiment than in Stone et al.'s (2001) study, whereas the SDs associated with these PSSs appear somewhat larger in the present study (21 msec for the same-position condition and 22 msec for the different-positions condition, in comparison with 8 msec in Stone et al.'s study; see their Table 1). The latter difference can probably be attributed to the smaller number of trials and SOAs used in our experiments than in Stone et al.'s study. The former difference might be "real," although we would hesitate to make any strong claims here given that a closer inspection of Stone et al.'s data reveals that it is only their Participants 16 and 17 (of a total of 17) who really stand out as having especially large SOAs.

For the fitted functions, the third defining parameter describes the peak probability of making a "simultaneous" judgment. These data were analyzed using the same method as for the PSS and SD data. In this case, a t test revealed significantly greater peak probabilities for the same-position data than for the different-positions data [$t(19) = 3.1, p = .005$]. This aspect of the data was further analyzed by carrying out a two-way within-participants analysis of variance (ANOVA) on the probability of "simultaneous" responses (for the raw data), with the factors of relative stimulus position (same vs. different) and
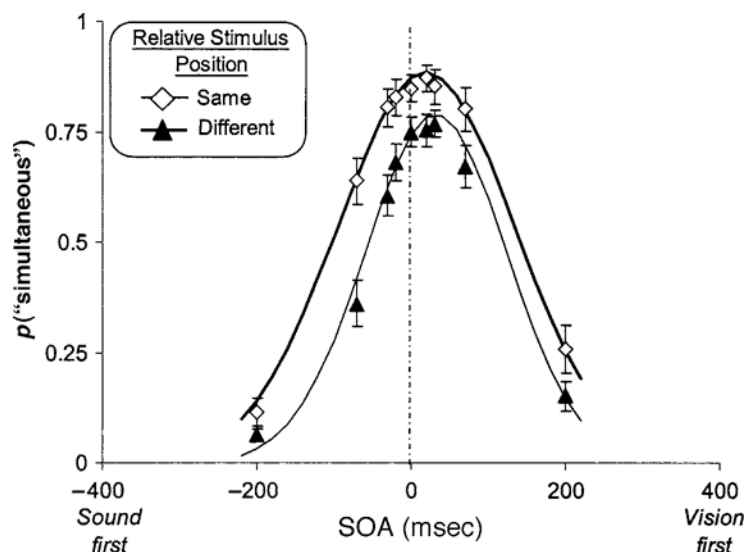
Figure 1. Mean proportion of "simultaneous" responses plotted as a function of the stimulus onset asynchrony (SOA) between the auditory and visual stimuli as a function of the relative stimulus position (same vs. different) in Experiment 1. Error bars represent the standard errors of the means for the grouped data. The average best-fitting functions are plotted for the same-position and different-positions data (see text for details). The results highlight the increased probability of participants' judging the two stimuli to be simultaneous when presented from the same spatial position rather than from different positions.

SOA ($-200, -70, -30 -20, 0, 20, 30, 70$, and 200 msec). This analysis revealed that the proportion of "simultaneous" responses varied as a function of the SOA [$F(8,12) = 50.6, p < .001, \eta^2 = .65$], as one would expect. There were also more "simultaneous" responses for the same-position condition than for the different-positions condition [$F(1,19) = 22.2, p < .001, \eta^2 = .05$]. Crucially, the effect of relative stimulus position was modulated by the SOA, leading to an interaction between these two factors [$F(8,12) = 3.6, p = .023, \eta^2 = .01$]. Pairwise comparisons between the same-position and different-positions conditions for each SOA revealed that the participants were significantly more likely to make a simultaneous response when stimuli were presented from the same spatial location at SOAs of $-70, -30, -20, 20$, and 200 msec (all $ps < .006$ after Bonferroni correcting for nine tests). Thus, the participants were more likely to make a "simultaneous" response when the stimuli were presented from the same position than when they were presented from different positions, specifically when the stimuli were nearly simultaneous (i.e., for SOAs around zero).

Discussion

The results of Experiment 1 demonstrate for the first time that judgments of the simultaneity (vs. successiveness) of pairs of auditory and visual stimuli are modulated by whether or not the stimuli are presented from the same spatial location: The participants in our study were more likely to report the two stimuli as being simultaneous when they were presented from the same spatial location than when they were presented from different spatial locations. Having demonstrated the role of spatial factors in audio–visual simultaneity perception, we went on to investigate whether the range of SOAs over which simultaneity perception is tested would also modulate the apparent perception of simultaneity. In particular, we wondered whether the *SD* of a participant's performance in the simultaneous/successive task actually reflects genuine limits on his or her perception of simultaneity, or whether instead it might be affected by the range of SOAs over which the perception of simultaneity was tested. That is, we thought it possible that decisional factors, judgment strategies, and/or any response biases adopted by participants might also modulate their apparent sensitivity to multisensory asynchrony. Any such finding would bring into question the appropriate interpretation of the results of a number of previous simultaneity studies (e.g., Hirsh & Fraisse, 1964; ITU-T, 1990; Michotte, 1912; Rihs, 1995).

Previously, researchers have noted that one important aspect of people's responses on certain psychophysical tasks is that they may exhibit a tendency to equalize the probability with which they make each of the responses that are available to them (see, e.g., Erlebacher & Sekuler, 1971; Parducci & Haugen, 1967; Sekuler & Erlebacher, 1971; see also Whipple, Sanford, & Colgrove, 1899, p. 283). If the participants had adopted such a *response equalization* strategy in the present study (or, for that matter, in any previous study in which a simultaneous-vs.-successive judgment task was used), then one might

Table 1
Results of Experiment 2: Mean Gaussian Curve Fits for the Simultaneity Judgment Task for Conditions in Which the Auditory and Visual Stimuli Were Presented From the Same Location Versus From Different Locations

	Same Location								Different Location								
	PSS		SD		Peak p				PSS		SD		Peak p				
Participant	M	SD	M	SD	M	SD	50% Range	−ln(MLE)	M	SD	M	SD	M	SD	50% Range	−ln(MLE)
1	18.5	13.7	80	16.9	.87	.03	−107 to 144	.43	20.44	18.7	79	22.2	.79	.05	−100 to 141	.51
2	−5.5	31.7	109	44.5	.86	.04	−176 to 165	.46	22.0*	7.3	99	5.0	.74†	.03	−123 to 167	.56
3	26.4	65.9	126	55.0	.66	.05	−149 to 202	.62	25.3	136.9	111	78.8	.47	.18	−99 to 150	.61
4	39.2	88.9	166	275.8	1.00	.01	−236 to 315	.25	52.7	31.5	110	34.5	.89†	.04	−123 to 228	.49
5	12.6	15.4	94	25.8	.95	.02	−140 to 166	.32	44.0*	17.2	62	15.0	.79†	.06	−50 to 138	.50
6	1.0	10.8	99	31.4	1.00	.00	−163 to 165	.12	30.1*	14.2	67	12.8	.89†	.05	−76 to 136	.41
7	12.1	37.0	154	86.4	.98	.01	−243 to 267	.25	14.7*	3.8	102	6.8	.97	.01	−153 to 182	.33
8	17.5	22.6	101	34.9	.93	.02	−146 to 181	.38	29.2*	14.7	77	12.9	.62†	.06	−75 to 133	.56
9	25.7*	7.6	91.5	10.2	.98	.01	−125 to 177	.29	53.9*†	8.4	73	6.0	.90†	.02	−62 to 170	.48
10	52.0*	10.1	88	7.3	.86	.03	−87 to 191	.51	52.4*	15.3	99	12.7	.75	.07	−94 to 199	.58
11	30.1	15.0	247.4	33.0	.94	.02	−373 to 433	.34	67.2†	24.4	142†	32.3	.63†	.05	−127 to 261	.64
12	37.5	3.6	118	5.0	.71	.02	−133 to 208	.60	51.0*	22.5	141	18.9	.58†	.03	−133 to 235	.65
13	4.1	17.6	87	14.9	.69	.06	−119 to 127	.56	23.3*	7.9	73	12.6	.94†	.02	−96 to 143	.35
14	44.2*	9.3	89	11.3	.94	.03	−101 to 189	.44	34.8*	12.5	91	11.5	.96	.02	−114 to 183	.38
15	7.0	5.7	35	4.0	.52	.04	−35 to 49	.44	19.9*	3.9	31	3.3	.61	.09	−22 to 62	.43
16	6.6	13.2	113	17.6	.96	.02	−178 to 191	.31	28.8*	10.4	93	12.3	.96	.03	−124 to 182	.34
17	25.8	15.6	100	56.6	.99	.02	−140 to 192	.27	32.8	24.4	73	14.9	.68†	.11	−71 to 137	.55
18	39.6*	15.9	180	51.8	1.00	.00	−261 to 340	.21	27.1	47.5	136	171.4	1.00	.00	−199 to 253	.22
19	22.6*	6.3	109	5.5	.99	.01	−158 to 204	.28	29.6*	11.4	72†	10.8	.84†	.03	−83 to 142	.47
20	−29.5	18.2	102	19.3	.87	.04	−190 to 131	.47	−16.8	4.5	95	5.7	.79	.02	−160 to 127	.52
Mean	19.4		114		.88				32.1		91		.79			

Note—The stimulus onset asynchronies (SOAs) used in this experiment ranged from −200 to +200 msec. PSS, point of subjective simultaneity (in milliseconds). Peak p, fitted probability of making a "simultaneous" response at the PSS. For the parameters PSS, peak p, and standard deviation (SD), bootstrap estimates of the SD are included (see, e.g., Efron & Tibishirani, 1993). 50% Range, range of SOAs over which each participant has a better than 50% chance of producing a "simultaneous" response. −ln(MLE), (negative) maximum likelihood estimate value for each of the fitted curves; the closer a value is to zero, the better the curve fit. Each −ln(MLE) value was obtained by weighting the log-likelihood value at each data point by the number of observations available at that point and subsequently renormalizing the overall likelihood by dividing by the total number of observations available for each function.   *Value significantly different from zero.   †Significant differences between parameters across the different stimulus positions.

predict that the apparent threshold for the perception of simultaneity derived from such studies would be dependent on the range of SOAs over which performance was assessed. We address this potentially important issue in the next experiment.

EXPERIMENT 2

In Experiment 2, the range of SOAs over which the perception of simultaneity was tested was twice as broad as that used in Experiment 1. The prediction was that participants might appear to be less sensitive to audio–visual asynchrony when tested over a wider range of SOAs than when tested over a narrower range (cf. Spence et al., 2003, p. 324, on this point). Such a result would suggest that the threshold of simultaneity derived from simultaneity studies is dependent on the range of SOAs tested. By contrast, if participants base their responses solely on their perceptual experience (as they were instructed to do), then the *SD*s derived from studies of simultaneity perception should be relatively independent of the range of SOAs tested (and might then remain similar to those reported in Experiment 1). Experiment 2 also allowed us to replicate the main effect of relative spatial position first outlined in the previous experiment, hence demonstrating the robustness of the effect.

Method

Twenty participants (mean age, 25 years) took part in the experiment as paid volunteers. All of the participants were right-handed by self-report. None of them had taken part in the previous experiment. The apparatus, materials, design, and procedure were identical to those used in Experiment 1, with the sole exception that all the SOAs were doubled from those reported in the previous experiment. SOAs of -400, -140, -60, -40, 0, 40, 60, 140, and 400 msec were used in this experiment.

Results and Discussion

As in the previous experiment, we used the Kolmogorov–Smirnov goodness-of-fit test to compare the observed distribution of responses with a normal distribution for each participant. All of the participants were considered in the subsequent analyses because all observations could reasonably have come from the specified distribution ($p > .05$; see note 3). The curve fit parameters for each participant are presented in Table 2, and the curves described by the means of these parameters are shown in Figure 2. Sixteen of the 20 participants required the visual stimulus to be presented further in advance of the auditory stimulus for the PSS to be achieved when the two stimuli were presented from different spatial locations ($M = 33.8$ msec) in comparison to when they came from the same spatial location ($M = 24.1$ msec). This difference was statistically significant at the group level [$t(19) = 3.0, p = .007$]. Seventeen of the 20 participants also showed larger *SD* values for the same-position condition than for the different-positions condition, a pattern of results that was also statistically significant at the group level [$t(19) = 4.2, p < .001$]. Finally, the peak proportion of "simultaneous" responses was significantly
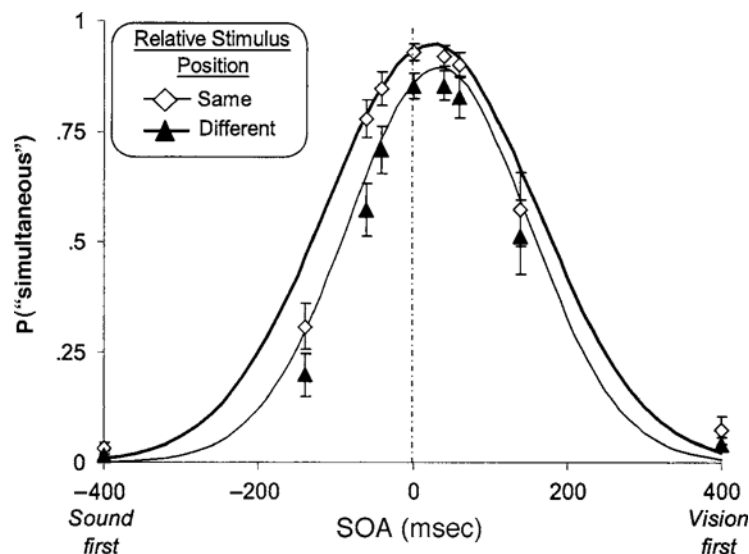

Figure 2. Mean proportion of "simultaneous" responses plotted as a function of the stimulus onset asynchrony (SOA) between the auditory and visual stimuli as a function of the relative stimulus position (same vs. different) in Experiment 2. The SOA values used in this experiment were all twice as large as those used in Experiment 1. All other experimental details remained the same. Error bars represent the standard errors of the means for the grouped data. The average best-fitting functions are plotted for the same-position and different-positions data (see text for details). The results highlight the increased probability of participants' judging the two stimuli to be simultaneous when presented from the same spatial positions rather than from different positions.

Table 2
Results of Experiment 2: Mean Gaussian Curve Fits for the Simultaneity Judgment Task for Conditions in Which the Auditory and Visual Stimuli Were Presented From the Same Location Versus From Different Locations

	Same Location								Different Location							
	PSS		SD		Peak p				PSS		SD		Peak p			
Participant	M	SD	M	SD	M	SD	50% Range	−ln(MLE)	M	SD	M	SD	M	SD	50% Range	−ln(MLE)
1	69.1*	21.5	174	23.9	.96	.03	−217 to 355	.39	67.3	35.7	168	39.2	.96	.03	−253 to 296	.38
2	21.5	32.6	172	22.4	1	.00	−265 to 308	.21	52.6*	10.5	146	12.1	.98	.01	−209 to 275	.30
3	22.9*	3.3	53	3.8	.75	.04	−56 to 102	.40	20.9	11.3	40	7.5	.73	.14	−55 to 61	.35
4	18.7	15.4	174	15.1	1	.02	−270 to 308	.26	35.5*	11.9	133†	13.8	1	.02	−206 to 237	.28
5	60.3*	26.7	244	157.9	1	.00	−347 to 467	.21	35.0*	17.5	202	69.8	1	.00	−309 to 363	.19
6	34.5	22.2	170	25.7	1	.00	−249 to 318	.16	71.5*	11.2	136	12.6	1	.01	−204 to 249	.31
7	12.0	9.3	158	18.1	.99	.01	−250 to 274	.26	28.5	35.3	107	33.5	.91	.10	−164 to 182	.43
8	9.7	12.7	270	35.1	1	.00	−440 to 459	.18	23.5	27.9	217	109.4	1	.00	−348 to 374	.19
9	0.4	12.6	110	14.0	.95	.03	−179 to 180	.34	4.2	13.3	111	14.9	.91	.03	−165 to 191	.39
10	30.5*	14.6	135	19.8	1	.01	−195 to 255	.22	43.1	23.8	156	33.3	.99	.00	−245 to 274	.21
11	31.6	19.9	130	21.5	.99	.03	−184 to 247	.33	35.7*	13.6	112	14.9	1	.01	−166 to 206	.30
12	−4.8	44.2	145	43.8	.85	.09	−231 to 221	.47	32.2*	16.3	133	12.6	.94	.04	−173 to 261	.36
13	31.4*	15.3	129	30.4	1	.00	−183 to 246	.18	45.8*	19.3	81	15.0	.89	.07	−114 to 145	.39
14	3.1	19.7	97	18.2	.89	.05	−152 to 158	.40	1.2	36.0	96	34.8	.53†	.08	−97 to 137	.50
15	42.6*	9.7	114	13.6	1	.01	−147 to 232	.24	45.2*	12.7	95	18.7	.97	.02	−146 to 166	.35
16	11.5	18.0	90	19.7	.91	.04	−134 to 157	.40	21.9	55.0	86	38.1	.68	.14	−103 to 139	.49
17	9.9	11.4	50	12.1	.80	.09	−66 to 86	.37	13.4*	4.1	36	4.4	.75	.12	−42 to 65	.33
18	29.1*	11.0	105	13.5	.94	.03	−141 to 200	.35	34.8*	14.6	122	26.4	1	.01	−192 to 214	.24
19	41.3	27.3	147	45.6	.99	.00	−202 to 285	.20	49.2*	9.4	106	9.2	.84†	.02	−137 to 192	.45
20	6.5	10.7	77	15.4	.95	.03	−120 to 133	.30	15.3	8.1	55	6.3	.85†	.05	−76 to 97	.38
Mean	24.1		137		0.95				33.8		117		0.90			

Note—The stimulus onset asynchronies used for this experiment ranged from −400 to +400 msec. See the note to Table 1 for additional explanations.

larger overall for the same-position data than for the different-positions data [$t(19) = 2.3$, $p = .035$], replicating the results of Experiment 1.

As in the first experiment, the data were further analyzed by means of a two-way within-participants ANOVA on the probability of "simultaneous" responses (for the raw data), with the factors of relative stimulus position (same vs. different) and SOA (-400, -140, -60 -40, 0, 40, 60, 140, or 400 msec). This analysis revealed that all effects were statistically significant [relative stimulus position, $F(1,19) = 20.6$, $p < .001$, $\eta^2 = .01$; SOA, $F(8,12) = 362.1$, $p < .001$, $\eta^2 = .73$; relative stimulus position $\times$ SOA, $F(8,12) = 4.3$, $p = .012$, $\eta^2 < .01$]. Subsequent Bonferroni-corrected t tests, used to compare across levels of relative stimulus position at each SOA, revealed significant differences at the SOAs of -140, -60, -40, and 40 msec.

The pattern of results across the two experiments was very similar. In order to assess this similarity more directly, we compared the results of the two experiments using separate experiment $\times$ relative stimulus position ANOVAs for each of the three curve fit parameters. In these analyses, relative stimulus position was a within-participants factor and experiment a between-participants factor. Analysis of the PSS values revealed a significant effect of relative stimulus position [$F(1,38) = 5.1$, $p = .031$] but no effect of experiment [$F(1,38) = 0.3$, $p = .567$] nor any interaction between these factors [$F(1,38) = 0.5$, $p = .502$]. For the SD data, there was an overall effect of relative stimulus position [$F(1,38) = 4.3$, $p = .045$] but no statistically significant effect of experiment [$F(1,38) = 3.1$, $p = .086$] nor any statistically significant interaction between these two factors [$F(1,38) = 0.1$, $p = .733$]. The third ANOVA, conducted on the fitted peak proportion of "simultaneous" responses, also provided a significant effect of relative stimulus position [$F(1,38) = 5.1$, $p = .029$]. There was also a statistically significant main effect of experiment [$F(1,38) = 5.9$, $p = .020$], with Experiment 1 showing lower overall peak proportions of "simultaneous" responses (.84) than Experiment 2 (.92). However, there was no interaction between these two factors [$F(1,38) = 1.2$, $p = .277$]. This set of between-experiments analyses therefore show that the effect of relative stimulus position is robust when tested across different SOA ranges. Moreover, the between-experiments analysis of the SD data, showing that the derived SD is independent of the range of SOAs (at least within the range that we tested), supports the idea the participants' performance was based on perceptual rather than decisional factors.

EXPERIMENT 3

Bertelson and Aschersleben (2003) have argued recently that the results observed in experiments in which the method of constant stimuli is used might be subject to postperceptual decision biases (see also Bertelson & de Gelder, 2004). They suggest that perceptual processes can be more successfully isolated through the use of psy-

chophysical staircase procedures (cf. Bertelson, 1999; Caclin, Soto-Faraco, Kingstone, & Spence, 2002; Welch, 1999a). Therefore, in Experiment 3 we used standard psychophysical parameter estimation by sequential testing (PEST) staircases (Taylor & Creelman, 1967) in order to investigate further whether or not the spatial modulation of simultaneity perception reflects a genuine perceptual interaction while once again attempting to rule out any decisional factors. The auditory and visual stimuli were now presented from the same position and from different positions in separate blocks of experimental trials (rather than interleaved on a trial-by-trial basis as in Experiments 1 and 2). Half of the participants were presented with the same-position blocks before the different-positions blocks, and the order was reversed for the remainder of the participants.

It could be argued that the participants in Experiments 1 and 2 might have judged the pairs of stimuli as being simultaneous more often when they were presented from the same spatial position because they may have picked up the fact that this was an important experimental variable from the very fact that the relative position of the stimuli (i.e., same vs. different) changed on a trial-by-trial basis. In our final experiment, we attempted to rule out any such potential influence on performance by blocking the order of presentation of same-position and different-positions trials. However, we thought that alternating the order of presentation of same-position and different-positions blocks of trials in this way might also affect the threshold of simultaneity. In particular, we thought it possible that when the first pairs of stimuli were presented from the same position, participants might be led to make an assumption of unity regarding the source of the audio–visual signals, which might result in a change of threshold for the subsequently presented different-positions trials (Welch, 1999a). By contrast, it seemed unlikely that perceptual thresholds would be affected by presentation of the stimuli from different positions first, because such presentation should not lead to any particular expectations regarding the unity of the auditory and visual stimuli.

Method

Participants. Twenty participants (mean age, 28 years) took part in Experiment 3. (Two additional participants were excluded from the study because their threshold values were greater than 400 msec, indicating their inability to perform the task.) All of the participants reported normal hearing and normal or corrected-to-normal vision. None had taken part in either of the two previous experiments. The apparatus and materials were identical to those used in the previous experiment. Once again, the participants had to judge the simultaneity of pairs of audio–visual stimuli, but now the stimulus presentation schedule was determined using the PEST staircase procedure (instead of the method of constant stimuli, as in Experiments 1 and 2).

Design. There were four within-participants factors: relative stimulus position (same vs. different), starting SOA for the staircase (0 vs. 500 msec), stimulus order (vision first vs. vision second; for staircases starting at the 0-msec SOA, stimulus order was determined by whether the visual event was presented first or second at the second step in the staircase), and side of the visual stimulus

Table 3
Means (*M*s) and Standard Errors (*SE*s) of Participants' Simultaneity Thresholds for Each of the 16 Staircases in Experiment 3

	Same Position								Different Position							
	Sound First				Vision First				Sound First				Vision First			
	Sound Left Vision Left		Sound Right Vision Right		Sound Left Vision Left		Sound Right Vision Right		Sound Left Vision Right		Sound Right Vision Left		Sound Left Vision Right		Sound Right Vision Left	
Starting SOA for Staircase	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
	Participants Who Performed the Same Position Staircases First															
0 msec	108	47	1,154	43	205	38	176	29	87	27	117	47	168	28	162	36
500 msec	86	40	94	26	113	26	81	25	73	22	77	29	95	25	119	29
	Participants Who Performed the Different Positions Staircases First															
0 msec	236	26	233	45	187	40	212	31	149	28	150	49	183	29	144	39
500 msec	105	42	192	27	179	27	208	27	119	23	123	31	150	26	114	31

Note—Sound First and Vision First indicate that the auditory event and the visual event, respectively, served as the first stimulus of the staircase.

(left vs. right). The same-position and different-positions conditions were presented in separate blocks of experimental trials. Three blocks of same-position trials were alternated with three blocks of different- positions trials, with the order of presentation counterbalanced across participants. At the beginning of the experiment, the participants were presented with a block of practice trials to acquaint them with the task (the relative stimulus position in the practice block was the same as that in the first block of experimental trials).

Procedure. The procedure was similar to that used in Experiments 1 and 2 with the following exceptions: In each block, the sequence of trials was controlled by eight randomly interleaved staircases. In all of the experimental blocks, half of the staircases began with an SOA of 500 msec and half with an SOA of 0 msec between the two stimuli. Each staircase was adjusted independently as a function of the participant's previous responses in that staircase, with the SOA shifted toward the point at which simultaneity and successive responses were equally frequent (50% threshold). The SOA was adjusted with a variable step size according to the PEST algorithm (Taylor & Creelman, 1967), with the initial step size set at 100 msec. A staircase was terminated when the step size fell below 10 msec in that staircase. The other staircases kept running until their terminations. In half of the experimental blocks, the stimuli were presented from the same position, and in the other half the stimuli were presented from different positions.

Results and Discussion

The averaged thresholds from each staircase (see Table 3) were submitted to a mixed between-participants/within-participants ANOVA with the factors of block order (same-position blocks first vs. same-position blocks second), stimulus order (vision first vs. vision second), relative stimulus position (same vs. different), first SOA of the staircase (0 vs. 500 msec), and side of the visual stimulus (left vs. right). For all of the analyses reported here, post hoc comparisons were made using Bonferroni-corrected *t* tests (where $p < .05$ prior to correction). The analysis revealed a significant main effect of relative stimulus position [$F(1,17) = 8.05$, $p = .011$], showing that the threshold for simultaneity was further from the 0-msec SOA when the auditory and visual stimuli were presented from the same position ($M = 161$ msec) than when they were presented from different positions ($M = 127$ msec). None of the other factors in the analysis was sig-

nificant. In particular, it is worth noting that the interaction between block order and relative stimulus position was not significant [$F(1,17) = 2.50$, $p = .132$], showing that the difference in performance between the same-position and different-positions conditions was consistent between the two groups of participants (for the group to which the same-position blocks were presented first, $M = 132$ msec for the same position and 110 msec for the different positions; for the group to which the same-position blocks were presented second, $M = 190$ msec for the same position and 141 msec for the different positions). We also investigated the possibility of any laterality effects in our data (i.e., any possible differences in performance between trials in which the sound was presented on the left and the light on the right or vice versa). We compared the conditions in which visual stimuli were presented on the left and auditory stimuli on the right with the conditions in which visual stimuli were presented on the right and auditory stimuli on the left. A *t*-test pairwise comparison revealed no significant difference between these two conditions [$t(19) = .245$, $p = .809$], suggesting that performance was not affected by the side from which each stimulus modality was presented on the different-sides trials.

The results of Experiment 3 therefore confirm the fact that participants' judgment of simultaneity for pairs of audio–visual stimuli is influenced by the relative positions from which the stimuli are presented. In particular, when the stimuli were presented from the same spatial location, the threshold of simultaneity was greater than when they were presented from different positions. In other words, subjective simultaneity was perceived across a wider range of SOAs when the stimuli were presented from the same location than when they were presented from different positions. Another finding to emerge from the analysis of Experiment 3 was that the order of presentation of the blocks of same-position and different-positions trials had no significant effect on the thresholds of simultaneity reported. In particular, when the different-positions blocks were presented after the same-position blocks, the threshold of simultaneity was no different from when the different-positions blocks were presented first. This

makes it less plausible that the participants' performance was dependent on the context in which the auditory and visual stimuli were presented (i.e., that their perception of simultaneity was affected by the order of presentation of the same-position and different-positions blocks of trials).

GENERAL DISCUSSION

The most important result to emerge from the analysis of Experiments 1 and 2 was the finding that people's judgments of whether or not two stimuli are simultaneous depends on whether or not the auditory and visual stimuli are presented from the same position. Judgments of simultaneity were more likely if the stimuli were presented from the same position rather than from different positions. Furthermore, analysis of the results of Experiment 3 revealed that the temporal window of subjective simultaneity was larger when the stimuli were presented from the same position than when they were presented from different positions. These findings have important implications for the interpretation of many previous studies of temporal perception (e.g., Hirsh & Fraisse, 1964; ITU-T, 1990; Michotte, 1912; Rihs, 1995; Van der Par et al., 1999) in which people's sensitivity to multisensory asynchrony may have been systematically overestimated by presentation of auditory and visual stimuli exclusively from different spatial locations (but see also Zampini, Shore, & Spence, 2003b). Previous studies have presumably failed to detect this spatiotemporal interaction in multisensory event perception because they focused on the effect of either just spatial or just temporal displacement on spatial ventriloquism effects, or on the effect of temporal asynchrony on simultaneity thresholds (see, e.g., Engel & Dougherty, 1971; Jack & Thurlow, 1973; Lewald et al., 2001; Radeau & Bertelson, 1987; Slutsky & Recanzone, 2001; see also Stone et al., 2001), but never on the effect of spatial displacement on temporal perception, as did the present study.

One reason for which the participants in our study were less inclined to respond "simultaneous" on trials in which the auditory and visual stimuli were presented from different positions than on those in which said stimuli were presented from the same position may be that they were using *apparent motion* cues to help them decide whether or not the stimuli were simultaneous (i.e., they may have been using any perception of apparent motion as a cue that the stimuli in different-positions trials must have been presented successively). Given that apparent motion cues could be expected to enhance performance only when the auditory and visual stimuli were presented from *different* spatial locations, a reliance on this cue would have resulted in a greater proportion of simultaneous responses in the same-location condition. However, researchers have argued that apparent motion does not occur when stimuli are presented to different sensory modalities (see, e.g., Allen & Kolers, 1981; Hirsh & Sherrick, 1961), despite clear evidence of the existence of apparent motion within both the auditory and visual modalities (see, e.g., Strybel, Manligas, Chan, & Perrott,

1990). It therefore seems unlikely that apparent motion cues can account for the increased probability of successive responses when the stimuli came from different spatial locations in the experiments reported here.

A more likely account of these results is that temporal resolution may be lost when separate sensory attributes of a particular object or event are bound together into a multisensory perceptual entity (see Spence et al., 2003). There is now mounting evidence of the existence of a cross-modal *temporal* ventriloquism effect (see, e.g., Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Scheier, Nijhawan, & Shimojo, 1999; Spence & Squire, 2003) similar to the more widely studied *spatial* ventriloquism effect (see, e.g., Bertelson & de Gelder, 2004, for a recent review). That is, stimuli from different sensory modalities presented at only small temporal asynchronies appear to be pulled together across time, resulting in the emergence of a unitary multisensory percept with a unique temporal onset. A similar pattern of results has also been demonstrated for the enhancement of the perception of audiovisual speech as a function of temporal asynchrony (see, e.g., Campbell & Dodd, 1980; Girin, Schwartz, & Feng, 2001; Massaro, Cohen, & Smeele, 1996; Navarra et al., 2005). Since multisensory integration and temporal ventriloquism are both more likely to occur when stimuli are presented from the same location than from different locations, this mechanism could account for the results observed here. Support for the "multisensory-integration-through-temporal-ventriloquism" account comes from the fact that the standard deviations (*SD*s) of the fitted functions were significantly larger for the same-position than for the different-positions condition. Therefore, it could be hypothesized that the participants exhibited poorer discriminative performance (larger *SD*s) in the same-position condition than in the different-positions condition because in the former condition the perception of simultaneity was stronger than in the latter, as the tendency to respond "simultaneous" more frequently might reveal (see the peak height of the distribution analysis).

The multisensory binding account (which is perceptual in nature) can be contrasted with a more decisional account of apparent intersensory interactions, which has been put forward to explain the *spatial* ventriloquism effect. According to Bertelson and Aschersleben (1998; see also Choe, Welch, Gilford, & Juola, 1975; Vroomen, 1999; Welch, 1999b; Welch & Warren, 1980), the spatial ventriloquism effects reported in many previous studies may have been caused, at least in part, by participants' voluntarily assuming that the auditory and visual stimuli ought to go together and then basing their responses on this assumption of unity. The extent to which people assume that a particular combination of auditory and visual stimuli should go together may then determine how they will respond in any particular experimental situation. Bertelson and Aschersleben (1998) argued that participants in previous spatial ventriloquism studies might have responded that auditory and visual stimuli came from the same location not because this is necessarily what they perceived, but instead because it is what they assumed

should have been the case from the context of the experiment in which they found themselves participating.

This argument could be extended to the simultaneity task studied here if one accepts that people are more likely to make an assumption of unity, and so to judge pairs of stimuli as being simultaneous, when they come from the same spatial location rather than from different spatial locations (see Engel & Dougherty, 1971; Fraisse, 1964; Miller, 1972; Piéron, 1952). In particular, discrete auditory and visual signals are more likely to be perceived as referring to a single multisensory event rather than to two separate events when they are presented in spatiotemporal proximity, thus inducing a sensation of cross-modal "phenomenal causality" (Guski & Troje, 2003). It could be argued that the participants in Experiments 1 and 2 may have found it more natural to respond "simultaneous" when the stimuli were presented from the same location than when they were presented from different locations. Any such assumption of unity for spatially colocalized audio–visual stimulus pairs may actually have been enhanced by the fact that the relative stimulus position factor (i.e., same position vs. different positions) varied randomly on a trial-by-trial basis in our first two experiments. Furthermore, certain researchers (e.g., Bertelson & Aschersleben, 2003; Bertelson & de Gelder, 2004) have also argued that the method of constant stimuli, as used in our first two experiments, may be more susceptible to any such potential postperceptual decision biases than psychophysical staircase procedures are.

We therefore conducted a final experiment (Experiment 3) using psychophysical PEST staircases (see Taylor & Creelman, 1967) instead of the method of constant stimuli. Furthermore, we blocked the presentation of the same-position and the different-positions trials in an attempt to make this aspect of our experimental design less salient to our participants. The presumption was that those participants who started the experiment with the same-position trials may have been more likely to develop an assumption regarding the unity of the audio–visual stimulus pairs, whereas those who started with the different-positions trials should have been less likely to do so (given that the stimuli were never colocalized in space for these participants in the first half of their experimental sessions). It is of crucial importance that we obtained exactly the same pattern of results in Experiment 3 as in our two earlier experiments. Furthermore, there was absolutely no effect of whether the stimuli were initially presented from the same position or from different positions. These results therefore confirm that participants' judgments of the simultaneity of audio–visual stimulus pairs are genuinely influenced by the relative position from which stimuli are presented (i.e., even when potential decisional accounts of our findings had been ruled out).

The comparison of the results of Experiments 1 and 2 was designed to clarify whether or not the *SD*s for simultaneity perception that were derived from the participants' performance were affected by decisional factors. If the participants were to adopt a response equalization strategy in our study (as has been suggested by some previous research; see Erlebacher & Sekuler, 1971; Parducci & Haugen, 1967; Sekuler & Erlebacher, 1971), then their performance (and hence the *SD*s calculated from their responses) should have been dependent on the range of SOAs tested. However, the results derived from the between-experiments comparison of the *SD*s support the perceptual account, since the threshold for the perception of simultaneity wasn't influenced by the range of SOAs over which performance was assessed (i.e., the analysis of *SD* values derived from Experiments 1 and 2 revealed that neither the main effect of experiment nor the interaction between experiment and relative stimulus position was significant).

It will be interesting in future experiments to investigate the role of spatial factors on temporal perception in more ecologically relevant situations (e.g., in the case of audiovisual speech perception). Our expectation is that the window for multisensory integration should be larger for more ecologically valid stimulus combinations, and there might possibly be less of a role for spatial factors given the increased spatial and temporal ventriloquism effects that are seen when ecologically valid multisensory stimuli are used (see, e.g., Dixon & Spitz, 1980; Navarra et al., 2005; Spence & Squire, 2003). It should be noted, though, that no study of the influence of increasing the ecological validity of the stimuli on multisensory integration, in which the possibility of response biases has been completely eliminated, has yet been conducted (see Bertelson & Aschersleben, 1998; Caclin et al., 2002).

Spatial Modulation of the PSS

Presentation of visual and auditory stimuli from the same spatial position results in a larger temporal window for simultaneity perception. As has been shown previously (see, e.g., Jaśkowski, Jaroszyk, & Hojan-Jesierska, 1990; Slutsky & Recanzone, 2001; Van de Par et al., 1999), visual stimuli had to be presented before auditory stimuli in order for the two to be perceived as simultaneous. Importantly, however, we found in the present study that the visual stimulus had to lead the auditory stimulus by a smaller time interval when the two stimuli were presented from the same spatial location than when they were presented from different locations. As we have already pointed out, this result may reflect the fact that the auditory and visual stimuli were more likely to be ventriloquized toward each other in time (hence reducing the effective PSS) when presented from the same spatial location than when presented from different spatial locations. This result is also consistent, however, with an attention-shifting account of temporal perception, for it has been argued by some researchers that people must shift their attention from the first stimulus to the second in order to judge whether or not two events are simultaneous (Allan, 1975; Kristofferson, 1967; but see Shore, Spry, & Spence, 2002). It will presumably take more time for attention to shift to the second stimulus if it is presented from a different spatial location than the first, since that will require not only a shift of attention from one modality to another (see, e.g., Spence,

Nicholls, & Driver, 2001) but also a shift of spatial attention from one location to another (see, e.g., Rhodes, 1987; Shulman, Remington, & McLean, 1979; Spence, 2001). Given that it presumably takes longer to shift attention from one modality to another *and* from one location to another than simply to shift attention between sensory modalities at the same spatial location, this might also explain why vision needs more lead time when the visual stimulus is presented from a position different from that of the sound stimulus. Whatever the underlying cause of this effect may be, the point remains that the PSS for pairs of auditory and visual stimuli is critically dependent on the relative spatial positioning of the stimuli, thus adding to the list of factors that have been shown to modulate the PSS (see, e.g., Jaśkowski, 1999).

Implications for Studies on the Neural Substrates of Multisensory Simultaneity Detection

Cognitive neuroscientists have recently started to investigate the neural correlates of the conscious perception of multisensory simultaneity using techniques such as positron emission tomography (Bushara, Grafman, & Hallett, 2001; see also Rao, Mayer, & Harrington, 2001) and event-related functional magnetic resonance imaging (Raizada & Poldrack, 2001). Using simultaneity tasks similar to the one described here, Bushara et al. have demonstrated that the detection of multisensory asynchrony (as opposed to synchrony) results in the activation of a large-scale dynamic network of cortical and subcortical areas, including the insula, the posterior parietal and prefrontal cortex, and cerebellar areas. The auditory and visual stimuli in these neuroimaging studies were, however, presented from different spatial positions: The auditory stimuli were presented over headphones, whereas the visual stimuli were presented in front of the participants. This spatial discrepancy may be particularly important when it comes to assigning a functional role to these networks of neural activation (see Macaluso, George, Dolan, Spence, & Driver, 2004, on this point). As we have stated already, the spatiotemporal conditions that give rise to the perception of simultaneity for pairs of auditory and visual stimuli in these studies are very similar to those that give rise to spatial ventriloquism effects (see Bertelson & de Gelder, 2004, for a review). It is unclear, therefore, to what extent the networks of activation reported in these neuroimaging studies are related specifically to multisensory synchrony detection per se rather than to the neural processes associated with spatial ventriloquism (see, e.g., Bertelson, 1998; Slutsky & Recanzone, 2001). Given the unknown contributions of spatial and temporal factors in these neuroimaging studies, the present findings highlight how important it is that future neuroimaging (and behavioral) studies recognize the consequences of the spatial confounds identified in this research in order to investigate the neural loci of specifically temporal multisensory binding (see Macaluso et al., 2004).

REFERENCES

ALLAN, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception & Psychophysics*, **18**, 29-36.

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

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.: Psychology Press.

BERTELSON, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347-362). Amsterdam: Elsevier.

BERTELSON, P., & ASCHERSLEBEN, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, **5**, 482-489.

BERTELSON, P., & ASCHERSLEBEN, G. (2003). Temporal ventriloquism. Crossmodal interaction on the time dimension: 1. Evidence from time order judgments. *International Journal of Psychophysiology*, **50**, 147-155.

BERTELSON, P., & DE GELDER, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141-177). Oxford: Oxford University Press.

BERTELSON, P., VROOMEN, J., & DE GELDER, B. (1997). Auditory–visual interaction in voice localization and in bimodal speech recognition: The effects of desynchronization. In C. Benoit & R. Campbell (Eds.), *Proceedings of the Workshop on Audio–Visual Speech Processing* (pp. 97-100). Rhodes.

BUSHARA, K. O., GRAFMAN, J., & HALLETT, M. (2001). Neural correlates of auditory–visual stimulus onset asynchrony detection. *Journal of Neuroscience*, **21**, 300-304.

CACLIN, A., SOTO-FARACO, S., KINGSTONE, A., & SPENCE, C. (2002). Tactile "capture" of audition. *Perception & Psychophysics*, **64**, 616-630.

CALVERT, G. A., BRAMMER, M. J., & IVERSEN, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, **2**, 247-253.

CAMPBELL, R., & DODD, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, **32**, 85-99.

CHOE, C. S., WELCH, R. B., GILFORD, R. M., & JUOLA, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, **18**, 55-60.

DIXON, N. N. F., & SPITZ, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719-721.

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

EFRON, B., & TIBSHIRANI, R. (1993). *An introduction to the bootstrap.* London: Chapman & Hall.

ENGEL, G. R., & DOUGHERTY, W. G. (1971). Visual–auditory distance constancy. *Nature*, **234**, 308.

ERLEBACHER, A., & SEKULER, R. (1971). Response frequency equalization: A bias model for psychophysics. *Perception & Psychophysics*, **9**, 315-320.

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

FRAISSE, P. (1964). *The psychology of time.* London: Eyre & Spottiswoode.

GIRIN, L., SCHWARTZ, J. L., & FENG, G. (2001) Audiovisual enhancement of speech in noise. *Journal of the Acoustical Society of America*, **109**, 3007-3020.

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

GUSKI, R., & TROJE, N. F. (2003). Audiovisual phenomenal causality. *Perception & Psychophysics*, **65**, 789–800.

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.

ITU-T (1990). *Television and sound transmission: Tolerances for transmission time differences between the vision and sound components of a television signal*. International Telecommunication Union, Telecommunication standardization sector of ITU, Recommendation J.100.

JACK, C. E., & THURLOW, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual & Motor Skills*, **37**, 967-979.

JAŚKOWSKI, P. (1999). Reaction time and temporal-order judgment as measures of perceptual latency: The problem of dissociations. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 265-282). Amsterdam: Elsevier, North-Holland.

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

JONES, J. A., & MUNHALL, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, **25**, 13-19.

KRISTOFFERSON, A. B. (1967). Attention and psychophysical time. *Acta Psychologica*, **27**, 93-100.

LEWALD, J., EHRENSTEIN, W. H., & GUSKI, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, **121**, 69-79.

MACALUSO, E., GEORGE, N., DOLAN, R., SPENCE, C., & DRIVER, J. (2004). Spatiotemporal contributions to audiovisual speech perception: A PET study. *NeuroImage*, **21**, 725-732.

MASSARO, D. W., COHEN, M. M., & SMEELE, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, **100**, 1777-1786.

MCGRATH, M., & SUMMERFIELD, Q. (1985). Intermodal timing relations and audio–visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, **77**, 678-685.

MICHOTTE, A. (1912). Nouvelles recherches sur la simultanéité apparente d'impressions disparates périodiques (expérience de complication) [New research on the apparent simultaneity of periodic heterogeneous stimuli (complication situation)]. *Annales de l'Institut Supérieur de Philosophie*, **1**, 571-663.

MILLER, E. A. (1972). Interactions of vision and touch in conflict and nonconflict form perception tasks. *Journal of Experimental Psychology*, **96**, 114-123.

MOLLON, J. D., & PERKINS, A. J. (1996). Errors of judgement at Greenwich in 1796. *Nature*, **380**, 101-102.

MOREIN-ZAMIR, S., SOTO-FARACO, S., & KINGSTONE, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, **17**, 154-163.

MUNHALL, K., & VATIKIOTIS-BATESON, E. (2004). Specialized spatio-temporal integration constraints of speech. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 177-188). Cambridge, MA: MIT Press.

MYUNG, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, **47**, 90-100.

NAVARRA, J., VATAKIS, A., ZAMPINI, M., SOTO-FARACO, S., HUMPHREYS, W., & SPENCE, C. (2005). *Exposure to asynchronous audiovisual speech increases the temporal window for audiovisual integration of non-speech stimuli*. Manuscript submitted for publication.

PANDEV, P. C., KUNOV, H., & ABEL, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *Journal of Auditory Research*, **26**, 27-41.

PARDUCCI, A., & HAUGEN, R. (1967). The frequency principle for comparative judgments. *Perception & Psychophysics*, **2**, 81-82.

PIÉRON, H. (1952). *The sensations: Their functions, processes and mechanisms*. London: Muller.

RADEAU, M., & BERTELSON, P. (1977). Adaptation to auditory–visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, **22**, 137-146.

RADEAU, M., & BERTELSON, P. (1987). Auditory–visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychological Research*, **49**, 17-22.

RAIZADA, R. D. S., & POLDRACK, R. A. (2001). Event-related fMRI of audio–visual simultaneity perception. *Journal of Cognitive Neuroscience*, **14** (Suppl.), 172.

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

RHODES, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics*, **42**, 1-14.

RIHS, S. (1995). The influence of audio on perceived picture quality and subjective audio–visual delay tolerance. In R. Hamberg & H. de Ridder (Eds.), *Proceedings of the MOSAIC workshop: Advanced methods for the evaluation of television picture quality* (pp. 133-137). Eindhoven.

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

SEKULER, R., & ERLEBACHER, A. (1971). The invalidity of "invalid results from the method of constant stimuli": A common artifact in the methods of psychophysics. *Perception & Psychophysics*, **9**, 309-311.

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

SHULMAN, G. L., REMINGTON, R. W., & MCLEAN, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 522-526.

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

SPENCE, C. (2001). Crossmodal attentional capture: A controversy resolved? In C. Folk & B. Gibson (Eds.), *Attention, distraction and action: Multiple perspectives on attentional capture* (pp. 231-262). Amsterdam: Elsevier.

SPENCE, C., BADDELEY, R., ZAMPINI, M., JAMES, R., & SHORE, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, **65**, 318-328.

SPENCE, C., & DRIVER, J. (EDS.) (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

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

SPENCE, C., & SQUIRE, S. B. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, **13**, R519-R521.

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.

SUGITA, Y., & SUZUKI, Y. (2003). Implicit estimation of sound-arrival time. *Nature*, **421**, 911.

TAYLOR, M. M., & CREELMAN, C. D. (1967). PEST: Efficiency estimates on probability functions. *Journal of the Acoustical Society of America*, **41**, 782-787.

VAN DE PAR, S., KOHLRAUSCH, A., & JUOLA, J. F. (1999, November). *Judged synchrony/asynchrony for light–tone pairs*. Poster presented at the 40th annual meeting of the Psychonomic Society, Los Angeles.

VAN DE PAR, S., KOHLRAUSCH, A., & JUOLA, J. F. (2004). *Synchrony judgments and temporal discrimination thresholds for audio–visual stimulus pairs*. Manuscript submitted for publication.

VROOMEN, J. (1999). Ventriloquism and the nature of the unity assumption. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 389-393). Amsterdam: Elsevier.

WELCH, R. B. (1999a). The advantages and limitations of the psychophysical staircase procedure in the study of intersensory bias: Commentary on Bertelson. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 363-369). Amsterdam: Elsevier.

WELCH, R. B. (1999b). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contribu-*

tions to the perception of spatial and temporal events (pp. 371-387). Amsterdam: Elsevier.

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**, 280-286.

ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003a). Audiovisual temporal order judgments. *Experimental Brain Research*, **152**, 198-210.

ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003b). Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology*, **50**, 165-180.

NOTES

1. The definition of the PSS used in the present study and in other simultaneity judgment studies (see, e.g., Stone et al., 2001) is different from that derived from TOJ studies. In audiovisual TOJ studies, the PSS typically represents the average interval by which one stimulus must lead the other in order for participants to make the "sound first" and "light first" responses with equal probability (see, e.g., Spence, Baddeley, Zampini, James, & Shore, 2003; Sugita & Suzuki, 2003; Zampini, Shore, & Spence, 2003a, 2003b). It has typically been assumed that the PSS derived from TOJ tasks (which really reflects the point of maximal uncertainty on the part of the participant regarding the temporal ordering of the stimuli) will provide an accurate measure of when the participant will be most likely to perceive the two stimuli as being simultaneous. However, this assumption has not, to our knowledge, been tested empirically. It will therefore be an interesting question for future research to determine whether the PSS derived from TOJ studies really does correspond to the PSS derived if participants are instead required to make a simultaneous/successive judgments (but see Van der Par, Kohlrausch, & Juola, 2004, on this issue).

2. In our own previous research on the role of relative spatial position (i.e., same vs. different) on TOJ performance, we have found results similar to those reported here for simultaneity judgments (Spence et al., 2003; Zampini et al., 2003a). In particular, better discriminative performance and larger PSSs were reported when stimuli are presented from different positions rather than from the same position (but see Zampini et al., 2003b). However, as has already been mentioned (see note 1), it is unclear whether the PSS calculated for a TOJ task provides a measure of subjective perceived simultaneity or just a measure of the point at which participants are most uncertain.

3. The Kolmogorov–Smirnov statistic, D, measures the largest difference between the empirical data and the test distribution. Therefore, nonsignificant values of D indicate that the empirical data may have been plausibly sampled from the test distribution.