The spatial constraint in intersensory pairing

Vroomen, J.; Keetels, M.N.

Published in:
Journal of Experimental Psychology. Human perception and performance

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

Journal of Experimental Psychology:
Human Perception and Performance
2006, Vol. 32, No. 4, 1063–1071

0096-1523/06/$12.00 DOI: 10.1037/0096-1523.32.4.1063

The Spatial Constraint in Intersensory Pairing: No Role in Temporal Ventriloquism

Jean Vroomen and Mirjam Keetels
Tilburg University

A sound presented in temporal proximity to a light can alter the perceived temporal occurrence of that light (temporal ventriloquism). The authors explored whether spatial discordance between the sound and light affects this phenomenon. Participants made temporal order judgments about which of 2 lights appeared first, while they heard sounds before the 1st and after the 2nd light. Sensitivity was higher (i.e., a lower just noticeable difference) when the sound–light interval was ~100 ms rather than ~0 ms. This temporal ventriloquist effect was unaffected by whether sounds came from the same or a different position as the lights, whether the sounds were static or moved, or whether they came from the same or opposite sides of fixation. Yet, discordant sounds interfered with speeded visual discrimination. These results challenge the view that intersensory interactions in general require spatial correspondence between the stimuli.

Keywords: intersensory perception, spatial discordance, temporal order judgment, temporal ventriloquism, spatial attention

The question of how sensory modalities cooperate to form a coherent representation of the environment is the focus of much behavioral and neuroscientific research (Calvert, Spence, & Stein, 2004). The most commonly held view among researchers in multisensory perception is what has been referred to as the *assumption of unity,* which states that as events from different modalities share more (amodal) properties, the more likely it will be that they originate from a common object or source (e.g., Bedford, 1989; Bertelson, 1999; Radeau, 1994; Stein & Meredith, 1993; Welch, 1999; Welch & Warren, 1980). Although there is some dispute as to what counts as a common property (e.g., Bertelson, 1999), without doubt the two most important ones are considered to be commonality in *space and time* (e.g., Radeau, 1994). Indeed, in the real world, signals in different modalities from a common object or event will often be spatially and temporally aligned, and it seems only logical that multisensory integration is constrained by these two factors. Following this notion, intersensory integration should be reduced or absent when stimuli are too far apart in space or time, because in that case two objects or events will be perceived rather than a single multimodal one.

Constraints on intersensory perception have mostly been studied with conflict situations where incongruent information about potentially the same distal event is presented to different modalities (de Gelder & Bertelson, 2004). A well-known example is the *ventriloquist effect* where the perceived location of target sounds is displaced toward light flashes delivered simultaneously at some distance, despite instructions to ignore the lights (for recent reviews, see de Gelder & Bertelson, 2003; Vroomen & de Gelder, 2004a). Several behavioral and physiological studies have shown that the ventriloquist effect disappears when the audiovisual synchrony exceeds approximately 300 ms or when the disparity between the sound and light exceeds 15 degrees (e.g., Godfroy, Roumes, & Dauchy, 2003; Hairston et al., 2003; Lewald & Guski, 2003; Radeau & Bertelson, 1977; Slutsky & Recanzone, 2001), although the specific degree of tolerated disparities ranges widely (e.g., Wallace et al., 2004).

Similar constraints on intersensory perception have been demonstrated at the neurophysiological level, in particular multisensory neurons in the superior colliculus. As shown by the work of Stein and coworkers (Stein & Meredith, 1993), a subset of these neurons shows the greatest increase in firing rates (compared to unimodal baselines) when a sound and a light are approximately synchronous and come from approximately the same position in space. At large spatial disparities, the auditory stimulus loses its capacity to enhance the effectiveness of the visual stimulus, and can actually become an inhibitor of its salience. These principles of multisensory integration are not only believed to underlie behavior where the superior colliculus is likely to be involved, like attentive and orientation behavior or saccadic reaction time (Frens & Van Opstal, 1998), but in recent years they have become the prime example of how intersensory integration in general is organized in the brain (Calvert et al., 2004).

Here, though, we will question whether the *spatial rule* should be considered as a *general* constraint on multisensory integration. Most of the studies that demonstrated spatial congruency effects on intersensory processing either required a spatial response (as in the spatial ventriloquist situation where participants localize a target sound), or it was likely that some form of spatial attention was critically involved. For example, it is known that participants are faster to detect or discriminate a light at an unknown location when a sound is presented briefly before the light at the same location rather than a different one (for recent reviews, see Driver

Jean Vroomen and Mirjam Keetels, Department of Psychology, Tilburg University, Tilburg, the Netherlands.

Correspondence concerning this article should be addressed to Jean Vroomen, Tilburg University, Department of Psychology, Warandelaan 2, PO Box 90153, 5000 LE, Tilburg, the Netherlands. E-mail: j.vroomen@uvt.nl

& Spence, 2004; Spence & McDonald, 2004). Some intersensory phenomena, though, like the well-known McGurk effect whereby a heard speech sound is affected by seen lip movements (McGurk & MacDonald, 1976), are almost exclusively affected by temporal synchrony rather than by the relative spatial location of the auditory and visual stimuli (Bertelson, Vroomen, Wiegeraad, & de Gelder, 1994; Colin, Radeau, Deltenre, & Morais, 2001). Temporal and spatial constraints in audiovisual speech processing are thus separable, and this has been confirmed in a PET study where violations of temporal and spatial constraints activated different brain regions (Macaluso, George, Dolan, Spence, & Driver, 2004). Other examples of intersensory phenomena that do not rely on spatial alignment were reported by Stein, London, Wilkinson, and Price (1996), who found that auditory stimuli alter judgments about perceived visual intensity of a light (with the position of the light known in advance) *regardless* of whether the sound was coincident with the visual stimulus or displaced by 45 degrees to the left or right (see also Vroomen & de Gelder, 2000). Similarly, Welch, DuttonHurt, and Warren (1986) reported that a fluttering sound altered the perceived rate of a flickering light, regardless of whether the sound and light were emanating from a single spatial locus (straight ahead) or whether they were separated by 45 or 90 degrees (see also Recanzone, 2003). Detection of infrequent deviants has been shown to be better for multisensory stimuli rather than unisensory ones (the redundant target effect), and this effect is, at the behavioral level, also rather independent of spatial alignment (Murray et al., 2004; Teder-Salejarvi, Di Russo, McDonald, & Hillyard, 2005). From those examples, it thus seems reasonable to assume that the spatial rule on intersensory integration may not apply if spatial attributes of the experimental task or spatial attention are not critically involved.

To examine this issue more systematically, we conducted a series of experiments on *temporal ventriloquism*, which is an intersensory illusion in the temporal domain. The basic phenomenon is that a sound in temporal proximity of a light can attract the temporal occurrence of the light (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Scheier, Nijhawan, & Shimojo, 1999; Stekelenburg & Vroomen, 2005; Vroomen & de Gelder, 2004b). As an example, Vroomen and de Gelder (2004b) studied this phenomenon using the flash-lag effect (FLE). The FLE is a visual illusion wherein a flash and a moving object that are presented in the same location are perceived to be displaced from one another. When the flash and the moving stimulus are physically aligned, observers typically report the flash as lagging behind the moving stimulus. The authors observed that when a noise burst was synchronized with the flash, it reduced the magnitude and the variability of the FLE, relative to a silent condition. Moreover, when the sound was presented before, at, or after the flash ($\pm\sim$100 ms), the size of the FLE varied linearly with the delay of the sound, thus indicating that an isolated sound sharpened the temporal boundaries of a flash, and when desynchronized with the flash, attracted its temporal occurrence in the order of about 5% of the temporal discrepancy.

Here we examined whether spatial discordance between a sound and light affects temporal ventriloquism. If the common notion on intersensory integration—spatial congruency always matters—is correct, one expects no or less temporal ventriloquism when there is a spatial mismatch between the auditory and visual inputs. If, in contrast, there is no such effect, it would demonstrate that the dominant view on intersensory integration needs to be qualified, and that it does not apply in the temporal domain. To explore this, we adopted a visual temporal order judgment task (TOJ) developed by Scheier et al. (1999) and Morein-Zamir et al. (2003). Participants were presented pairs of visual stimuli at various stimulus onset asynchronies (SOAs) and were asked to judge which of two lights (upper or lower) appeared first. Task-irrelevant sounds were presented before the first and after the second light so that, due to temporal ventriloquism, the lights would seem to occur farther apart in time, thereby improving sensitivity (i.e., lowering the Just Noticeable Difference, JND). The sounds themselves were completely orthogonal to the task at hand, and did not predict in any sense which light appeared first. Response conflicts between the position of the sounds and the lights could therefore be excluded.

In a pilot study not reported here ($N = 14$), we determined the optimal sound–light interval for obtaining a temporal ventriloquist effect by varying the interval from ~0 ms to ~400 ms in steps of ~100 ms. Results showed that, compared to the ~0 ms condition, the effect was biggest when the audiovisual interval was ~100 ms, and that it became smaller when the interval was further increased. This fits well with the results of Morein-Zamir et al. (2003) who also found that the temporal ventriloquist effect was biggest around the ~100 ms interval. The ~100 ms interval was therefore chosen as the condition in which temporal ventriloquism should be at maximum, and it was compared to the ~0 ms interval that served as baseline. Note that in this comparison two sounds were presented in both conditions. This allowed to measure the effect of audiovisual temporal discordance (the difference between a ~100 ms vs. a ~0 ms interval) in its purest form, while other more general effects that sounds might have on visual processing (facilitatory or inhibitory) were subtracted.

Experiment 1

Two lights were vertically aligned in the median plane while sounds were presented at eye level either centrally, or far to the left or right at −90 or + 90 degrees (see Figure 1). Participants judged whether the upper or lower light was presented first, ignoring the sounds. The orientation of the lights (up vs. down) was orthogonal to the direction of the sounds (left, right, or middle) so that response conflicts between the sounds and lights were excluded. If spatial correspondence between sound and light were critical for intersensory integration to occur, one would expect more temporal ventriloquism (i.e., a larger improvement at the ~100 ms interval if compared to the ~0 ms interval) if sounds were presented centrally rather than laterally.

Method

Participants. Twenty students from Tilburg University received course credits for their participation.

Stimuli. Two red LEDs (diameter of 0.5 cm, luminance of 40 cd/m2) were positioned vertically at 5 degrees above or below central fixation, with fixation at eye level and at 90 cm distance. Sounds were presented via either one of two speakers. One speaker was positioned in between the two lights, the other was placed either at the far left, or at the far right (at 90 degrees), also at eye-level and at 90 cm distance. The auditory stimuli consisted of a short white noise burst of 5 ms presented at 70 dB(A). A

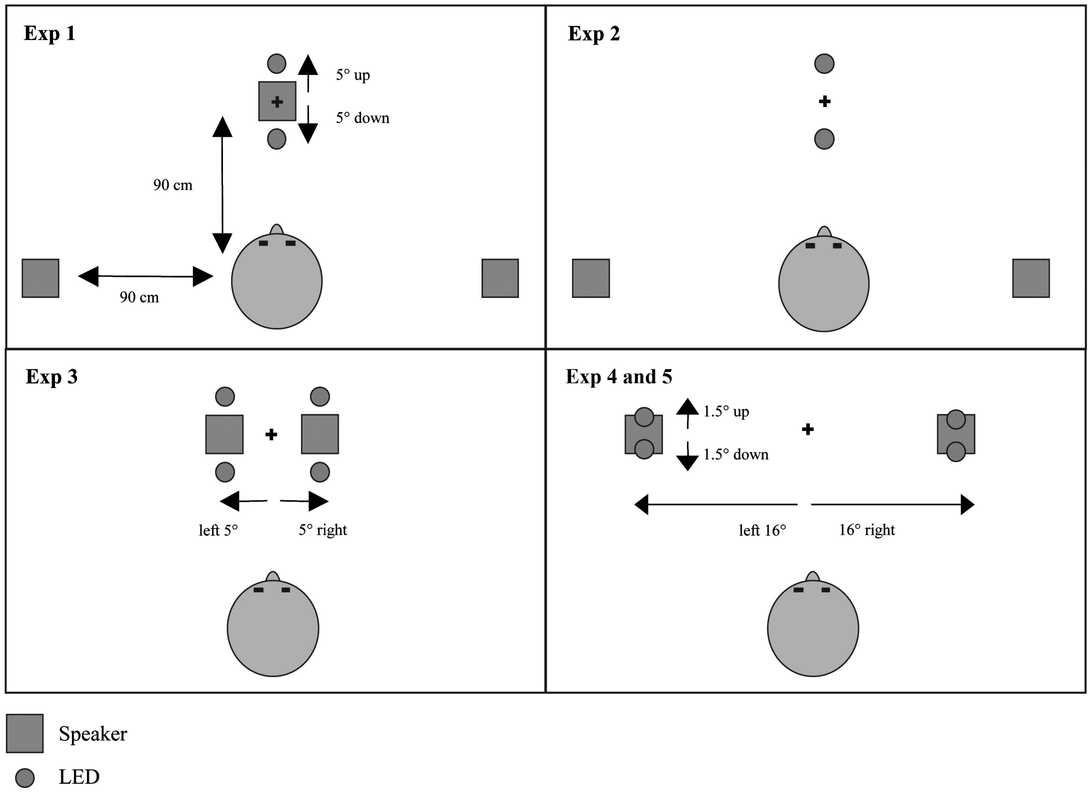

Figure 1. Schematic top view of the setup in Experiments 1–5.

small green LED on the center of the middle loudspeaker served as fixation.

Design. Three within-subjects factors were used: Sound location (Same or different azimuth as the lights), Audiovisual interval (~0 or ~100 ms), and SOA between the two visual stimuli (ranging from −75 ms to + 75 ms in steps of 15 ms, with negative values indicating that the lower LED was presented first). The audiovisual interval represents the time between the onset of the first sound and first light, and the time between the onset of the second light and the second sound. Thus, in the ~100 ms condition, the first sound preceded the first light by ~100 ms and the second sound trailed the second light by ~100 ms, while in the ~0 ms condition sounds and lights were synchronous. A silent visual-only condition was added to test whether the presence of a sound as such helped or interfered with performance. This resulted in 50 unique trials, each randomly presented 20 times in 4 blocks of 250 trials each. The side at which the lateral loudspeaker was placed varied between blocks. For half of the participants, the lateral speaker was positioned on the right in the first two blocks, and on the left in the other two blocks, whereas the order was reversed for the other half of the participants.

Procedure. Participants sat at a table in a dimly lit and soundproof booth. The fixation light was illuminated at the beginning of the experiment, and participants were instructed to maintain fixation on this central green LED during testing. Trials consisted of the onset of the first LED, and after the SOA, the second LED was turned on. Both LEDs remained lit until a response was made. The two sounds, if present, were presented according to the sound–light interval condition. The participants' task was to judge whether the lower or the upper LED was presented first by pressing one of two designated keys with the right thumb or index finger, respectively. Responses were unspeeded and participants were informed that the sounds should be ignored, as they did not predict in any way which light came first. The next trial started 2000 ms after a response was detected.

To acquaint participants with the TOJ task, experimental blocks were preceded by a practice session. The practice block consisted of 30 trials in which each combination of audiovisual interval and sound location at the ± 45, ± 60 and ± 75 ms SOAs was presented once. During practice, participants received verbal feedback ("Correct" or "Wrong") about whether they had pressed the correct key or not.

Results and Discussion

Trials of the practice session were excluded from analyses. The proportion of *up-first* responses was calculated for each combination of condition and SOA for each participant. To compute JNDs, proportions were converted into equivalent Z-scores assuming a cumulative normal distribution (Finney, 1964). For each condition, the best-fitting straight line was calculated over the 10 SOAs. These lines' slopes and intercepts were used to determine the JND (JND = 0.675/slope) and the point of subjective simultaneity (PSS). The JND represents the smallest interval between the onsets of the two lights needed for participants to be able to judge correctly which stimulus had been presented first on 75% of the trials. The PSS represents the average interval by which the upper light has to lead the lower one for being perceived as simultaneous.

In the 2 (Sound location) × 2 (Audiovisual interval) ANOVA on the PSSs, no effect was significant (Sound location, $F(1, 19) = 3.54$, $p = .075$; Audiovisual interval, $F < 1$; and the interaction, $F(1, 19) = 2.09$, $p = .16$; average PSS = +0.8 ms). This was as predicted, because there was no reason to expect that either one of the two lights (upper or lower) would be differentially affected by the sounds, thus creating an overall shift toward up or down responses. In the ANOVA on the JNDs (see Table 1), there was an

Table 1
Experiments 1–4: Mean Just Noticeable Differences (JND) in ms and Standard Errors of the Mean

		Sound from same location			Sound from different location		
	Visual only	0 ms	100 ms	TVE	0 ms	100 ms	TVE
Experiment 1	32.7 (3.4)	26.2 (1.7)	21.2 (1.5)	5.0	30.1 (2.8)	22.1 (1.4)	8.0
Experiment 2	41.3 (3.5)	33.5 (2.8)	24.9 (1.7)	8.6	38.1 (3.0)	29.5 (2.4)	8.6
Experiment 3	29.2 (2.8)	24.3 (1.6)	19.9 (0.8)	4.4	24.9 (1.7)	20.1 (0.7)	4.8
Experiment 4	28.2 (2.6)	32.3 (2.4)	27.7 (1.8)	4.6	32.8 (2.9)	27.5 (2.7)	5.3

Note. Standard errors of the mean in parentheses. The temporal ventriloquist effect (TVE) is the difference between the ~0 and ~100 ms intervals; ms = milliseconds

overall effect of the audiovisual interval, $F(1, 19) = 16.08$, $p < .001$, as JNDs in the ~100 ms interval were lower (i.e., better performance) than in the ~0 ms interval (a 6.5 ms overall temporal ventriloquist effect). There was also an overall effect of sound location, $F(1, 19) = 7.36$, $p < .025$, because performance was less accurate when sounds were presented from lateral positions rather than the central one (a 2.4 ms difference). Most important, the interaction between sound location and audiovisual interval did not approach significance, $F(1, 19) = 1.69$, $p = .21$, indicating that the temporal ventriloquist effect was essentially the same for sounds presented from the same versus different positions as the lights. Separate *t* tests comparing the ~0 ms versus ~100 ms intervals indeed confirmed that the 5.0 ms ventriloquist effect for sounds presented from the same position, $t(19) = 3.13$, $p < .005$, and the 8.0 ms effect for sounds presented from different positions, $t(19) = 3.45$, $p < .005$, were both significant. Note also that, if anything, temporal ventriloquism was actually bigger when sounds were presented from a different position rather than from the same position as the lights. A comparison between the central synchronous condition and the silent condition also showed that the presence of a sound enhanced rather than interfered with performance, $t(19) = 2.35$, $p < .05$.

The results of Experiment 1 thus essentially showed that there was no sign that temporal ventriloquism was attenuated when sounds were presented from a different position rather than from the same position as the lights. In both cases, JNDs improved about equally. This indicates that spatial colocalization is *not* a critical factor for audiovisual temporal ventriloquism to occur. Another relevant finding was that JNDs in general were worse when sounds were presented from a different position as the lights rather than the same. One plausible explanation for this overall disturbing effect of laterally presented sounds is that they distracted attention away from the target lights (Spence, Nicholls, Gillespie, & Driver, 1998), thus interfering with performance (see also Experiment 5). Whatever the interpretation of this effect, it shows that the incongruent location of the two sounds could not be totally ignored. Sound location was thus noticeable and affected JNDs; it just did not affect temporal ventriloquism. In the following experiments, we explored more variants of audiovisual spatial discordance.

Experiment 2

In Experiment 2, we compared *moving* versus static sounds. The two accessory sounds were presented either centrally (i.e., from a static position), or the first sound was presented from the far left, and the second from the far right (or vice versa). The latter gave the impression that the sounds moved from left to right (or right to left). The question was whether temporal ventriloquism would be affected by such a clear difference in spatial attributes of the sounds (static–central vs. moving–lateral).

Method

Participants. Sixteen new students participated.

Stimuli and design. All experimental details were as in Experiment 1, except that the two loudspeakers were placed one at each side of the participant's head (at 90 degrees, eye level, and at 90-cm distance, see Figure 1). For the central sound location, a stereo sound was presented from both loudspeakers so that the two sounds appeared to occur from the middle. For the moving sounds, in half of the trials the first (equally loud) sound was presented from the right speaker and the second from the left; for the other half of the trials the order was reversed. The position of the first sound (central, left sound first, or right sound first) varied randomly.

Results and Discussion

JNDs and PSSs were computed as before. A 2 (Sound location) × 2 (Audiovisual interval) ANOVA on the PSSs showed that there was a main effect of sound location, $F(1, 15) = 8.01$, $p < .025$, a marginally significant effect of audiovisual interval, $F(1, 15) = 4.32$, $p = .055$, and no interaction, $F(1, 15) = 2.94$, $p = .11$. The effect of sound location indicated that there were more up responses for moving sounds than central ones, a finding for which we have no clear explanation. More important, in the ANOVA on the JNDs the effect of audiovisual interval was again significant, $F(1, 15) = 18.30$, $p < .001$, as performance was more accurate in the ~100 ms interval than in the ~0 ms interval (an 8.6 ms overall temporal ventriloquist effect). There was also a main effect of sound location, $F(1, 15) = 8.50$, $p < .025$, because moving sounds distracted overall performance (a 4.6 ms effect) compared to static sounds. Most important, the interaction between sound location and the audiovisual interval was again *nonsignificant,* $F < 1$, indicating that the temporal ventriloquist effect was virtually identical for static versus moving sounds. Separate *t* tests confirmed that the 8.6 ms effect for static sounds, $t(15) = 3.48$, $p < .005$, and the 8.6 ms effect for moving sounds, $t(15) = 3.39$, $p < .005$, were both significant. Clearly, then, even auditory apparent lateral motion did not attenuate intersensory integration of centrally presented lights.

In the following two experiments, we tested yet another variant of audiovisual spatial discordance by checking whether it matters if sounds and lights are presented from the same or different sides of central fixation.

Experiment 3

The setup of Experiment 3 was changed so that the sounds and lights were presented either to the left or right of central fixation (see Figure 1). This allowed measurement of temporal ventriloquism for audiovisual stimuli whose components were presented from either the same or the opposite sides of the median. There were two reasons for presenting stimuli this way. First, this setup is very similar to the one used in the cross-modal orthogonal cueing paradigm introduced by Spence and Driver (Spence & Driver, 1997, 1994), where it has been shown that sound location is highly relevant for speeded visual target discrimination. In one of their conditions, spatially nonpredictive auditory cues were presented from either the left or right of a central fixation point, while after a variable SOA, visual targets were presented from one of four locations directly above or below the cue on either side. A robust finding from that literature is that the speeded up-down discrimination of the visual target is improved (i.e., faster reaction times) when the auditory cue is presented from the same side as the light at short SOAs (e.g., Experiment 1 of Spence & Driver, 1997). The location of the sound is thus highly relevant in this speeded detection task, presumably because it reflexively shifts spatial attention toward the target light. The question here was whether the location of the sound—and spatial attention—would be relevant for temporal ventriloquism. It has been argued that spatial attention modulates intersensory binding, as indicated by the fact that the neural response enhancement of audiovisual stimuli is bigger when stimuli are attended rather than unattended (Talsma & Woldorff, 2005). If this is of relevance at the behavioral level, one could expect more temporal ventriloquism if sounds and lights were presented from the same side rather than from opposite sides of the median.

The second reason for presenting the audiovisual stimuli at different sides of the median is that it has been shown that for *audiovisual* temporal order, it matters whether the sound and light cross the median or not. In an audiovisual TOJ task, participants are presented a sound and light at various SOAs and are asked to judge which came first (sound first or light first). Performance is better (i.e., lower JNDs) when sounds and lights are presented at different sides of fixation (Keetels & Vroomen, 2005; Zampini, Shore, & Spence, 2003a, 2003b) rather than the same side. One interpretation of this effect is that there is less intersensory binding (or less *fusion*) whenever stimuli are presented to the different hemispheres (Zampini et al., 2003a, 2003b). This so-called hemispheric account would thus also predict less temporal ventriloquism for sounds and lights presented at different sides of fixation, because the stimuli were processed by different hemispheres. Three accounts (the assumption of unity, spatial attention, and the hemispheric account) thus all predict less temporal ventriloquism if sounds and lights are presented at different sides of the median rather than the same.

Method

Participants. Twenty new students participated.

Stimuli. Stimuli were as before, except that the setup now consisted of 4 LEDs. Two LEDs were vertically aligned on the left at 5 degrees from central fixation, the other two LEDs were placed in symmetric positions on the right (see Figure 1). The two loudspeakers where positioned at eye level in between the vertically aligned LEDs, with their centers also at 5 degrees to the left or right. The horizontal separation of spatially incongruent sound–light stimuli was thus 10 degrees. A centrally positioned LED served as fixation.

Design. There were four within-subjects factors: Side of the light (Left or right), Sound location (Same or different side as the lights), Audiovisual interval (~0 ms or ~100 ms), and SOA between the two visual stimuli (ranging from −75 ms to + 75 ms in steps of 15 ms, with negative values indicating that the lower LED was presented first). A silent condition (side of light left or right) was added to test whether a sound helped or interfered with performance. This resulted in 100 unique trials, all randomly presented 10 times in 5 blocks of 200 trials each. Participants were instructed to keep their eyes on fixation, and to judge whether one of the lower or one of the upper LEDs was presented first—ignoring side—by pressing one of two designated keys with their right thumb or index finger, respectively.

Results and Discussion

JNDs and PSSs were computed as before. A 2 (Side of light left or right) × 2 (Sound at same or different side as light) × 2 (Audiovisual interval ~0 or ~100 ms) ANOVA on the PSSs showed that no effect was significant (all Fs < 1). The same ANOVA on the JNDs showed that the Side of the light (left or right) had no overall effect, $F < 1$, nor did it interact with Location, $F(1, 19) = 1.64$, $p = .22$; or with Audiovisual interval, $F < 1$. The data were therefore pooled over side of light. In the 2 (Sound location same or different as light) × 2 (Audiovisual interval ~0 or ~100 ms) ANOVA on the JNDs, there was main effect of audio-visual interval, $F(1, 19) = 14.32$, $p < .001$, as performance was more accurate in the ~100 ms interval than in the ~0 ms interval (a 4.6 ms overall temporal ventriloquist effect). Sound location (Same or different side) had no overall effect, $F < 1$, nor was the theoretically important interaction between the sound–light interval and sound location significant, $F < 1$. Separate t tests confirmed that the 4.4 ms effect for sounds presented from the same side, $t(19) = 3.35$, $p < .005$ and the 4.8 ms effect for sounds presented form the opposite side, $t(19) = 3.37$, $p < .005$, were both significant.

Results thus showed that there were *equal* amounts of temporal ventriloquism for sounds presented from the same versus the opposite side of the lights. The prediction drawn from the assumption of unity, the spatial attention account, and the hemispheric account was thus not supported by the data, indicating that audiovisual spatial correspondence plays a much less critical role in intersensory pairing than might be expected on the basis of these theories. In the following experiment, we further explored this important observation.

Experiment 4

Experiment 3 showed that for temporal ventriloquism it did not matter whether sounds were presented from the same or from the opposite side of the lights. Somewhat surprising, though, there also was no overall disturbing effect of the incongruent sound location on the JNDs, while this effect was present in Experiments 1 and 2. A possible reason for this might be that the spatial separation of the incongruent sound–light pairs was relatively large in Experiments 1 and 2 (i.e., 90 degrees separation), while this was much smaller (i.e., 10 degrees) in Experiment 3. The value of 10 degrees was chosen because we considered that the TOJ task might become too difficult when the lights were presented too far out in the visual periphery. To check, though, whether participants actually noticed the audiovisual spatial discordance, we conducted another experiment in which the difference between the spatially congruent and incongruent trials was further increased.

Method

Participants. Sixteen new students participated.

Stimuli and design. Stimuli and design were as in Experiment 3, except that the horizontal separation of the spatially incongruent sound–light stimuli was increased from 10 degrees to 32 degrees (see Figure 1), while the vertical separation of the congruent sound–light stimuli was further decreased from 5 degrees to 1.5 degrees. This manipulation should further increase the difference between the spatially congruent and incongruent trials.

Results and Discussion

The data were analyzed as in Experiment 3. In the 2 (Sound location at same or different side as light) $\times$ 2 (Audio-visual interval ~0 or ~100 ms) ANOVA on the PSSs, no effect was significant (all Fs < 1). The same ANOVA on the JNDs showed that there was a main effect of audiovisual interval, $F(1, 15) = 10.20$, $p < .01$, as performance was more accurate in the ~100 ms interval than in the ~0 ms interval (a 4.9 ms overall temporal ventriloquist effect). Sound location (Same or different side) had again no overall effect, $F < 1$, nor was the theoretically important interaction between the sound–light interval and sound location significant, $F < 1$. Separate t tests showed that the 4.6 ms ventriloquist effect for sounds presented from the same side was marginally significant, $t(15) = 2.11$, $p = .051$, while the 5.3 ms ventriloquist effect for sounds presented from the opposite side was significant, $t(15) = 2.48$, $p < .025$. The results thus again showed that temporal ventriloquism was unaffected by whether sounds were presented from the same or the opposite side of the lights. Yet, despite that the horizontal separation of the incongruent stimuli was further increased, there was again no overall effect of sound location on the JNDs.

Between-Experiments Analysis

We also conducted an omnibus ANOVA to check whether the size of temporal ventriloquist effect of the Experiments 1 – 4 were different. A 2 (Sound at same or different location as light) $\times$ 2 (Audiovisual interval ~0 or ~100 ms) ANOVA with Experiment (1–4) as a between-subjects factor showed that there was a main effect of Experiment $F(3, 68) = 6.57$, $p = .001$, as the overall JNDs were somewhat worse in Experiment 2. There was also a main effect of Sound location, $F(1, 68) = 10.11$, $p < .005$, that interacted with Experiment, $F(3, 68) = 2.88$, $p < .05$, indicating that spatially discordant sounds interfered with overall performance in Experiments 1 and 2, but not so in Experiments 3 and 4. There was, furthermore, a highly significant main effect of Audiovisual interval, $F(1, 68) = 59.12$, $p < .001$, (i.e., the temporal ventriloquist effect) that, importantly, did not interact with Experiment, $F(3, 68) = 1.23$, $p = .30$, thus indicating that the size of the temporal ventriloquist effect was the same across experiments. Finally, and most important, there was no interaction between Sound location and Audiovisual interval, and no second-order interaction between Sound location, Audiovisual interval, and Experiment (both Fs < 1), showing that, essentially, in all four experiments temporal ventriloquism was unaffected by audiovisual spatial discordance.

Experiment 5

The between-experiments comparison showed that while in Experiments 1 and 2 there was an overall effect of sound location on the JNDs, this was not the case in Experiments 3 and 4. This leaves the possibility that participants in Experiments 3 and 4 were able to effectively ignore sound location. As a further control on this, we changed the TOJ task into a speeded discrimination task. By stressing speed, the task becomes very similar to the cross-modal orthogonal cueing task of Spence and Driver (1997) where it has been shown that sound location is relevant for visual target discrimination. This final control experiment allowed us to check whether the absence of an effect of sound location in the TOJ task of Experiments 3 and 4 was due to either the specific setup used here, or rather that the effect of sound location—and presumably reflexive spatial attention—is task-dependent. If indeed an effect of sound location were found in speeded discrimination while not in temporal order judgment, it would indicate that temporal ventriloquism can be dissociated from where spatial attention is attracted.

The task of the participants in Experiment 5 was to make speeded discriminations about which of two lights (upper or lower) appeared first, ignoring their side (left or right). The stimuli were as in Experiment 4, except that the SOA between the two lights was kept at + 75 ms or −75 ms. This was an interval for which it could be relatively easy decided which of the two lights appeared first, so that speed rather than accuracy could be emphasized. The two sounds were presented as before, at either ~0 or ~100 ms intervals. It should be noted that the stimuli of the ~100 ms interval are like the ones used in the cross-modal orthogonal cueing task of Spence and Driver (1997, Experiment 1 with auditory cues preceding visual targets at 100 ms), the main difference being that their stimuli consisted of a single sound (the cue) and a single flash (the target) rather than two sounds and two light onsets. Spence and Driver (1997) also did not use the ~0 ms cue–target interval. The relevant finding of their experiment was that participants were faster when a sound was presented from the same side rather than the opposite side of a visual target at ~100 ms SOA (a 33 ms spatial cueing effect). The question here was whether similar effects of sound location would show up in our setup if we used their task.

Method

Participants. Thirteen new students participated.

Stimuli and design. The setup was as in Experiment 4, but to emphasize speed while keeping errors within limits, we only used the + 75 ms or −75 ms SOA between the two lights. The two sounds were presented at either ~0 or ~100 ms intervals before the first and after the second light at either the same or the different side. This resulted in four unique conditions, each presented 80 times in randomized order (20 for each of the four possible LEDs) in two blocks of 160 trials each. A short practice session preceded the experimental trials. Reaction times were measured from the onset of the first light.

Results and Discussion

Incorrect responses and reaction times (RTs) above 1000 ms were discarded from the RT analysis. The intersubject means of participants' mean RTs and the corresponding error rates are shown in Table 2. In the 2 (Sound location at same or different side as light) × 2 (Audiovisual interval ~0 or ~100 ms) ANOVA on the RTs, there was a main effect of sound location, $F(1, 12) = 11.85$, $p < .005$, because RTs were faster when the sound was at the same side as the visual target rather than the opposite side (a 15 ms overall spatial cueing effect). RTs were also faster in the ~100 ms interval rather than the ~0 ms interval, $F(1, 12) = 8.14$, $p < .025$, while the interaction between sound location and sound–light interval was not significant, $F(1, 12) = 2.68$, $p = .12$.

The error rates were analyzed in the same way as the RTs. In the 2 (Sound location at same or different side as light) × 2 (Audiovisual interval ~0 or ~100 ms) ANOVA on the error rates, only the interaction was significant, $F(1, 12) = 9.15$, $p < .025$. Inspection of the table shows that performance was most accurate when the first sound preceded the visual target from the same position as the light at the ~100 ms sound–light interval. Separate *t* tests confirmed that less errors were made at the ~100 ms sound–light interval when the sound came from the same position as the target rather than the other side, $t(12) = 2.24$, $p < .05$, while there was no difference at the ~0 ms sound–light interval.

The results of Experiment 5 thus showed that there were clear effects of sound location on speeded discrimination, as participants were faster and made fewer errors when the sound came from the same position as the visual target. These results are in close correspondence with those of Spence and Driver (1997, Experiment 1), and thus show that it is not the specific setup used here, but the task that determines whether effects of spatial correspondence and reflexive spatial attention emerge.

General Discussion

We explored whether spatial correspondence between a sound and light affects temporal ventriloquism. Participants' judgments about which of two lights appeared first were more sensitive (compared to synchronized sounds) when a sound was presented before the first and after the second light, thus demonstrating temporal ventriloquism (Morein-Zamir et al., 2003; Scheier et al., 1999; Vroomen & de Gelder, 2004b). In four experiments, we observed that spatial discordance between the sounds and lights had no effect whatsoever on this phenomenon. Thus, whether the sounds came from the same medial position as the lights, or at 90 degrees to the left or right (Experiment 1), whether the sounds were presented statically or moved laterally (Experiment 2), or whether the sounds were presented from the same or the opposite side of the median (Experiments 3 and 4), in all cases temporal ventriloquism was unaffected. Moreover, Experiment 5 showed that spatially incongruent sounds nevertheless attracted spatial attention. Taken together, these results challenge the view that intersensory integration in general requires commonality in space or that it is strongly dependent on spatial attention.

Table 2
Experiment 5: Mean Reaction Times (RTs) and Percentages of Errors for Speeded Visual Target Discrimination

	Sound location			
	Same		Different	
Sound-light interval	RT	%	RT	%
0 ms	608	15	617	13
100 ms	580	12	602	15

Note. ms = milliseconds

A possible concern about these data might be that this interpretation is mainly based on null-effects of sound location, and that the manipulations might not have been powerful enough, or that the method might not have been sensitive enough to measure any effect of sound location. This seems unlikely, though, for several reasons. First, it should be noted that in all experiments we *did* obtain a temporal ventriloquist effect (lower JNDs at ~100 ms sound–light interval rather than ~0 ms) thus indicating that we were actually measuring in the sensitive range (i.e., an improvement was possible). Second, it is unlikely that sound location was not perceived or could be totally ignored, and that for this reason it had no effect on performance, as in Experiments 1 and 2 incongruent sounds did harm overall sensitivity, but not temporal ventriloquism. Third, while there was no overall disturbing effect of incongruent sounds on JNDs in Experiments 3 and 4 (sounds and lights to the left or right of fixation), there was nevertheless an effect on reaction times when the task was changed into speeded discrimination. The latter result is important as it demonstrates that sounds were effectively attracting spatial attention, thereby showing that temporal ventriloquism may not be strongly depending on spatial attention.

Another potential concern is related to the question whether the temporal ventriloquist effect as shown here is actually distinct from a more general warning or altering effect that sounds might have on visual processing (e.g., Posner, 1978). One possibility is that the first sound acted as a warning signal, thus alerting participants that visual stimuli were about to be presented. Such a warning signal might improve TOJ performance more at ~100 ms SOA than at ~0 ms SOA, which might imply that the observed JND improvements were not related to temporal ventriloquism as such, but rather to a more general attention phenomenon. Several observations, though, provide evidence against this possibility. As shown by Morein-Zamir et al. (2003, Experiments 2 and 3), when the first sound is presented either before or simultaneous with the first light, and the second sound either simultaneous with or after the second light, it is mainly the second sound presented *after* the second light that contributes to the effect. This thus essentially refutes the *warning* interpretation. Moreover, a warning account will not explain why sounds may sometimes interfere with performance if the sounds intervene in between the two lights. Performance then actually gets worse, presumably because the lights are now pulled together (Morein-Zamir et al., Experiment 4; Scheier et al., 1999). Finally, a last observation that is difficult to fit with the warning account is that a single sound presented either before the first light, after the second light, or in between the two lights, has no effect on TOJ performance (Morein-Zamir et al., Experiment 4; Scheier et al., 1999). This indicates—as would be expected by intersensory integration proper—that the two lights must be paired with two sounds.

The results also are relevant to the issue whether intersensory integration is a preattentive phenomenon or not. Several authors have argued that intersensory integration requires attention, as it has for example been observed that the amount of intersensory bias gets smaller when limited attentional resources are devoted to a secondary task (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; but see Tiippana, Andersen, & Sams, 2004; Vroomen, Driver, & de Gelder, 2001). Effects of spatial congruency on multisensory processing also have been observed in electrophysiological measures. For example, intersensory integration effects in event-related potentials (ERPs) tend to be larger in amplitude for attended rather than unattended stimuli (Talsma & Woldorff, 2005), although there is very little if any effect at the behavioral level (Murray et al., 2004; Teder-Salejarvi et al., 2005). The present results, though, seem to contrast with the idea that spatial attention is important for intersensory integration. As shown here, spatially discordant sounds were effectively attracting spatial attention away from the target lights, yet there was no effect on temporal ventriloquism. These results are therefore more in line with studies on spatial ventriloquism (i.e., the apparent location of a sound is shifted toward a spatially incongruent but synchronized flash) where it has been shown that intersensory integration occurs in a more automatic fashion, independent of where spatial attention is focused or attracted (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000; Bertelson, Vroomen, de Gelder, & Driver, 2000; Driver, 1996; Stekelenburg, Vroomen, & de Gelder, 2004; Vroomen, Bertelson, & de Gelder, 2001a, 2001b; Vroomen et al., 2001c). Clearly, the extent to which intersensory integration depends on spatial attention or attentional resources and its neurophysiological consequences is an issue that needs further investigation.

To conclude, here we showed that intersensory integration in the form of temporal ventriloquism can occur despite substantial spatial discordance between the individual components of the cross-modal stimulus. This demonstrates that the spatial rule about intersensory integration does *not* constitute a general constraint on intersensory pairing. Rather, it seems more likely that spatial discordance is only relevant to the extent that spatial attributes of the experimental task or spatial attention are critically involved. At first sight, this might seem counterintuitive because after all, most natural multisensory events are spatially and temporally aligned, except for some minor variations in time or space that people are readily able to adjust (e.g., Vroomen, Keetels, de Gelder, & Bertelson, 2004). However, a critical assumption that underlies the idea of the spatial rule is that spatial information has the same function in vision and audition. This notion, though, is arguable as it has been proposed that the role of space in hearing is only to steer vision (Heffner & Heffner, 1992), while in vision it is an indispensable attribute (Kubovy & Van Valkenburg, 2001). If one accepts that space has different functions in these modalities, then there is also no a priori reason why they should have the same value for intersensory interactions to occur.

References

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15,* 839–843.

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Cross-modal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50,* 157–163.

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology-Human Perception and Performance, 15,* 232–248.

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347–363). Amsterdam: North-Holland/Elsevier.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Cross-modal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50,* 147–155.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia, 38,* 1634–1642.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62,* 321–332.

Bertelson, P., Vroomen, J., Wiegeraad, G., & de Gelder, B. (1994). *Exploring the relation between McGurk interference and ventriloquism.* Paper presented at the International Congress on Spoken Language Processing, Yokohama, (pp. 559–562).

Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes.* Cambridge, MA: MIT Press.

Colin, C., Radeau, M., Deltenre, P., & Morais, J. (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica, 41,* 131–144.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7,* 460–467.

de Gelder, B., & Bertelson, P. (2004). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 8,* 7–7.

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

Driver, J., & Spence, C. (2004). Crossmodal links in endogenous spatial attention. In J. Driver (Ed.), *Crossmodal space and crossmodal attention* (pp. 179–220). Oxford, England: Oxford University Press.

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.* Cambridge, England: Cambridge University Press.

Frens, M. A., & Van Opstal, A. J. (1998). Visual auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Research Bulletin, 46,* 211–224.

Godfroy, M., Roumes, C., & Dauchy, P. (2003). Spatial variations of visual-auditory fusion areas. *Perception, 32,* 1233–1245.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience, 15,* 20–29.

Heffner, R. S., & Heffner, H. E. (1992). Visual factors in sound localization in mammals. *Journal of Comparative Neurolology, 317,* 219–232.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgements. *Experimental Brain Research, 167,* 635–640.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97–126.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16,* 468–478.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *Neuroimage, 21,* 725–732.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage, 21,* 125–135.

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

Radeau, M. (1994). Auditory-Visual Spatial Interaction and Modularity. *Cahiers de Psychologie Cognitive-Current Psychology of Cognition, 13,* 3–51.

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semi-realistic situations. *Perception & Psychophysics, 22,* 137–146.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology, 89,* 1078–1093.

Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science, 40,* 4169.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12,* 7–10.

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

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

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. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 555–574.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8,* 497–506.

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

Stekelenburg, J. J., & Vroomen, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport, 16,* 641–644.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters, 357,* 163–166.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17,* 1098–1114.

Teder-Salejarvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17,* 1396–1409.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology, 16,* 475–472.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica, 108,* 21–33.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63,* 651–659.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1583–1590.

Vroomen, J., & de Gelder, B. (2004a). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 141–150). Cambridge, MA: MIT Press.

Vroomen, J., & de Gelder, B. (2004b). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology-Human Perception and Performance, 30,* 513–518.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience, 1,* 382–387.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22,* 32–35.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158,* 252–258.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam: North-Holland/Elsevier.

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, 88,* 638–667.

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.

Received June 6, 2005
Revision received January 12, 2006
Accepted January 23, 2006 ■