Perception & Psychophysics
2002, 64 (4), 616-630

Tactile "capture" of audition

ANNE CACLIN
University of Oxford, Oxford, England
École Normale Supérieure, Paris, France

SALVADOR SOTO-FARACO
University of Oxford, Oxford, England
University of British Columbia, Vancouver, British Columbia, Canada

ALAN KINGSTONE
University of British Columbia, Vancouver, British Columbia, Canada

and

CHARLES SPENCE
University of Oxford, Oxford, England

Previous research has demonstrated that the localization of auditory or tactile stimuli can be biased by the simultaneous presentation of a visual stimulus from a different spatial position. We investigated whether auditory localization judgments could also be affected by the presentation of spatially displaced tactile stimuli, using a procedure designed to reveal perceptual interactions across modalities. Participants made left–right discrimination responses regarding the perceived location of sounds, which were presented either in isolation or together with tactile stimulation to the fingertips. The results demonstrate that the apparent location of a sound can be biased toward tactile stimulation when it is synchronous, but not when it is asynchronous, with the auditory event. Directing attention to the tactile modality did not increase the bias of sound localization toward synchronous tactile stimulation. These results provide the first demonstration of the tactile *capture* of audition.

Crossmodal interactions between audition and touch are the subject of increasing interest in the burgeoning literature on multisensory integration and attention in humans, in both the spatial and nonspatial domains (e.g., Eimer, Cockburn, Smedley, & Driver, 2001; Jousmäki & Hari, 1998; Merat, Spence, Withington, & McGlone, 1999; Sherrick, 1976; Spence, Nicholls, Gillespie, & Driver, 1998). For example, in the *nonspatial* domain, Jousmäki and Hari recently reported an audiotactile illusion whereby sounds synchronized with hand rubbing were shown to modify tactile sensations (e.g., enhancing the high frequencies of the hand-rubbing sound made the hands feel drier), an effect they labeled the *parchment-skin illusion.* In the *spatial* domain, Merat and colleagues have demonstrated that speeded elevation responses to auditory targets presented above or below the midline are influenced by simultaneous tactile distractors presented to the thumb (down) or index finger (up). Taken together, these studies suggest that extensive audiotactile interactions affect multisensory information processing. The present study was designed to explore audiotactile spatial interactions—specifically, whether perceived location of sounds would be affected by concurrent tactile stimulation presented to the fingertips.

Multisensory spatial interactions have often been shown through studies of intersensory bias. One example is the ventriloquism illusion: When simultaneous visual and auditory stimuli are presented from discrepant spatial positions, people tend to localize the sound as coming from a position closer to the visual event than is actually the case (e.g., Urbantschitsch, 1880; see Bertelson, 1998, for a recent review). The reverse effect, the biasing of visual localization by an auditory stimulus, has also been reported under certain conditions, although this effect is smaller in magnitude and less consistent (Radeau & Bertelson, 1976, 1987). Intersensory biases have also been shown to occur between other pairs of modalities: vision and touch (the so-called visual capture[1] of touch), vision and proprioception (e.g., Botvinick & Cohen, 1998; Pavani, Spence, & Driver, 2000; Pick, Warren, & Hay, 1969), and also audition and proprioception (e.g., Fisher, cited in Howard & Templeton, 1966; Pick et al., 1969).

The study of bias in sound localization as a function of tactile stimulation investigated here is important because neither modality has been shown to exhibit a clear dominance over the other in terms of spatial localization, unlike

A.C. was supported by the École Normale Supérieure de Paris and the Réseau de Sciences Cognitives d'Ile-de-France. S.S.-F. was supported by a network grant from the McDonnell-Pew Centre for Cognitive Neuroscience in Oxford and by a postdoctoral fellowship from the Killam Trust. Correspondence should be addressed to S. Soto-Faraco, Department of Psychology, 2136 West Mall, Vancouver, BC, V6T 1Z4 Canada (e-mail: ssoto@interchange.ubc.ca).

the more commonly studied situations, in which vision is one of the modalities presented (see Bertelson, 1998; Welch & Warren, 1986, for reviews). Vision has been described as the most accurate of the spatial senses (see Welch & Warren, 1986) because localization tasks are performed more precisely by vision than by any other sensory modality. There is also evidence that visual maps predominate over auditory or somatosensory maps with respect to spatial localization (Stein & Meredith, 1993; though see Auerbach & Sperling, 1974). At the behavioral level, Warren (1970) has demonstrated that auditory localization can be improved by the presentation of a structured visual environment, and this could be interpreted in terms of visual mapping of auditory stimuli. Additionally, neurophysiological findings in the monkey superior colliculus, a subcortical structure involved in multisensory integration (e.g., Stein & Meredith, 1993; Wallace, Wilkinson, & Stein, 1996) also support the hypothesis that visual maps have a unique and central role in both auditory (Jay & Sparks, 1984) and somatosensory (Groh & Sparks, 1996) spatial processing.

Pick et al. (1969) reported one of the few studies of multisensory spatial interactions that did not involve vision as one of the stimulus modalities. They introduced an artificial discrepancy between audition and proprioception by means of a pseudophone that displaced the azimuthal position of sounds by 11° laterally. With their vision occluded, participants rested a finger from their left hand on a loudspeaker placed above a table so that they could feel its operation whenever a series of auditory clicks were presented. Participants were informed explicitly that the auditory and proprioceptive stimuli might appear to come from discrepant locations and were required to point with their other hand under the table to indicate the perceived location of their finger and/or the sounds. Proprioception was shown to have a strong biasing effect on auditory judgments, whereas auditory stimuli exerted only a minimal influence on proprioceptive judgments. However, although Pick et al. described their study solely in terms of the "proprioceptive bias of audition," it is difficult to rule out the possibility that participants' judgments of the sound location may have been biased by where they felt the loudspeaker to be (i.e., by tactile cues regarding the loudspeaker location) rather than simply by where they felt their finger to be in space (i.e., by proprioceptive cues regarding their felt hand position), and hence Pick et al.'s results may reflect some unknown contribution of the tactile bias of audition.

Additional, albeit indirect, evidence for spatial interactions between audition and touch comes from a study reported by Freedman and Wilson (1967). They examined whether or not movement was a necessary prerequisite for compensation to auditory–tactile rearrangement. During the exposure phase of their experiment, participants listened to the sound of a moving loudspeaker cone that sometimes touched their fingers. As auditory inputs were artificially displaced laterally (again using a pseudophone), an audiotactile spatial conflict was created. Freedman and Wilson reported a shift in pointing responses to auditory targets following this exposure phase. This result may suggest that auditory and tactile inputs were integrated during the exposure period to overcome the induced spatial discrepancy. It is important to note, however, that the results of both Pick et al.'s (1969) study and Freedman and Wilson's study do not necessarily imply any integration of auditory and tactile information. During the exposure phase, participants were clearly aware that what they would feel at their fingertips would be the loudspeaker, and hence it could be argued that participants were using a *proprioceptive* cue regarding the loudspeaker's location to recalibrate auditory maps (cf. Pavani et al., 2000).[2]

Two different paradigms have been used so far to study intersensory bias: immediate effects and aftereffects. Typically, these paradigms reveal a shift in pointing responses either during the presentation of spatially discrepant stimuli in the two modalities (*immediate effects*) or after a period of exposure to an intermodal spatial conflict (*aftereffects* or *adaptation*). However, Choe and colleagues (Choe, Welch, Gilford, & Juola, 1975) claimed that immediate effects may reflect response biases rather than true perceptual interactions. That is, in many experiments using immediate effects, previous knowledge about the stimuli or just the obvious presence of conflicting cues may have induced response biases that can explain the intersensory effects found without the need to invoke perceptual processes (see Bertelson, 1998). Indeed, even the study of aftereffects does not seem to be completely free of potential cognitive biases. As Bertelson and Aschersleben (1998) have noted, the conscious detection of a spatial discrepancy during the exposure phase of an experiment could affect the setting of response criteria during the posttest period as well (as in Freedman & Wilson's, 1967, study). In support of this claim, Radeau and Bertelson (1974) reported that the magnitude of audiovisual aftereffects were affected by the particular instructional set given—that is, whether participants were told that the combination of lights and sounds had a common origin, a different origin, or else were given no information regarding the spatial relationship between the stimuli. Canon (1970, 1971) also demonstrated that the magnitude of aftereffects observed after an intermodal conflict were modulated by the particular modality attended during the exposure phase. Given that the study of genuine perceptual intersensory biases is relevant to understanding how we build up a coherent representation of external space (Bertelson & Aschersleben, 1998), it is clearly important to have a method that, as far as possible, avoids response biases in the study of intersensory bias.

Postperceptual adjustments, such as response bias, are more likely to occur when the situation is transparent (i.e., the data necessary for an explicit deliberation, such as the spatial separation of the inputs or their timing, can be perceived consciously by the participant; Bertelson, 1998), as in the majority of previous studies of intersensory bias. Bertelson and Aschersleben (1998) recently designed an experimental procedure free from such postperceptual adjustments to investigate audiovisual ventriloquism effects (see also Welch, 1999). This method uses the psychophysical staircase principle (Cornsweet, 1962) and provides a measure of any immediate perceptual effects. In Bertelson and Aschersleben's study, participants had to make a left–right discrimination (forced-choice response) regarding the position of sounds, whose locations were chosen using a psychophysical staircase procedure. They selected the sounds according to two randomly intermingled staircases, each staircase starting at the outermost right or left position, and moving according to the participant's responses (i.e., after a correct response in the left staircase, the next sound on the same staircase was moved to the right, and vice versa). This resulted in the two staircases converging toward a median location. Bertelson and Aschersleben reported that when a central flashing light was presented synchronously with the sound, response reversals (i.e., responses inconsistent with the response to the preceding trial belonging to the same staircase) occurred further from the center than when either the central flashing light was asynchronous with the sounds or else the light did not flash at all. Bertelson and Aschersleben argued that this result demonstrates that perception of the sounds was displaced toward the synchronized flashing light.

This psychophysical staircase procedure therefore reveals the genuine perceptual component of any multisensory spatial interaction, free from postperceptual adjustments elicited by the task or the experimental display. In the present study, we adapted Bertelson and Aschersleben's (1998) method to study audiotactile interactions. Irrelevant vibrotactile stimulation was always presented from directly in front of participants, who were required to make left– right discriminations regarding the location of the sounds. Consequently, felt touches could not induce or prime either of the two potential responses (left or right) to the target sounds. As in Bertelson and Aschersleben's study, after a given point in the staircases the participant was unaware of the actual spatial discrepancy between the stimuli, and the nontransparency condition was met (see Bertelson, 1998; Bertelson & Aschersleben, 1998).

EXPERIMENT 1

In our first experiment, we replicated Bertelson and Aschersleben's (1998) procedure except that we provided vibrations to the (centrally located) fingertips rather than visual stimuli. Any attraction of the perceived sound location toward the position of the tactile stimulation should manifest itself in the occurrence of incongruent responses further from the center (and thus earlier in the block) with synchronized vibrations presented to the fingertips relative to a no-vibration control condition.

Method

Participants. Eight participants (5 female), all right-handed by self-report, were tested in two sessions of about 20 min each. The mean age of participants was 23 ($SD = 2$) years. All participants reported normal hearing and touch and were naive as to the purpose of the experiment.

Apparatus and Materials. Participants were seated in a dark room with their chin in a chinrest, facing a black curtain. Two loudspeaker cones placed 46 cm apart (center to center) were hidden behind the curtain, at ear level (Figure 1). The loudspeakers were placed to the right (+30°) and left (−30°) of central fixation. Participants placed their index fingers on two vibrators (miniature linear actuators, TDITAC 8×12 mm; Transdimension Corporation, Irvine, CA) situated next to each other at the same height as the loudspeaker cones and directly between them. A cushion placed under the participants' elbows ensured a comfortable posture. The vibrators were placed inside a sponge cube into which the participants inserted their index fingers in order to attenuate any noise produced by the operation of the vibrators. A red LED, placed behind the curtain in front of the participant at the same height and distance as the loudspeakers, remained illuminated throughout the experiment to provide a central visual fixation point.

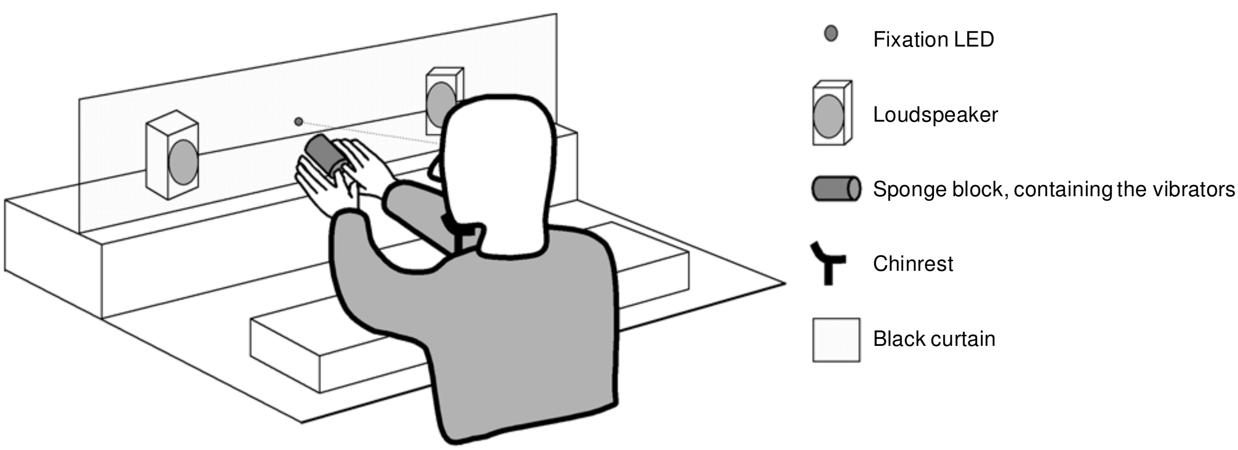

Figure 1. Schematic view of the participant and apparatus for testing the tactile capture of audition, as seen from a raised position behind the participant.

A personal computer running the Expe6 programming language (Pallier, Dupoux, & Jeannin, 1997) was used to control the presentation of stimuli and to register the participants' responses. Auditory stimuli were presented at different simulated locations ranging between the positions of the two loudspeakers. Two different types of stereophonic cues were used to control the perceived azimuthal position of the sound: interaural time differences (as in Bertelson & Aschersleben, 1998) and interaural intensity differences. Each type of stereophonic cue was tested separately. The reason for including two different auditory manipulations was to determine which cue type provided the most convincing evidence for the tactile capture of audition.[3] All sound files (stereo files with a sample rate of 22050 Hz) were generated with the CoolEdit 96 sound editor (Syntrillium Software Corporation, Scottsdale, AZ). The auditory stimuli consisted of three 15-msec, 2000-Hz sine-wave audible pulses (with rising and decaying linear envelopes, 2.5-msec rise-fall times), separated by 800 msec of silence (the total stimulus presentation lasted for 1,645 msec).

In the *time-difference* staircase, the temporal offset between the two channels ranged from -725 μsec to $+725$ μsec in steps of 45.35 μsec (corresponding to an angle of approximately 2.5º–3º), with the difference of 0 μsec corresponding to the center of the staircase. This staircase therefore provided 33 possible sound locations; 16 to the left, 16 to the right, and one located centrally. The *intensity-difference* staircase also provided 33 possible locations: 16 to each side plus a centrally located one. The extreme points of the staircases corresponded to one of the loudspeakers being turned on and the other turned off completely. At each step, 1/16 of the total sound pressure level (SPL) would be subtracted from one loudspeaker and added to the other loudspeaker (thus displacing the location of the sound one step toward the louder loudspeaker). The central point of the intensity staircase corresponds to sounds in which the intensity is equal for both loudspeaker cones. Note that with this procedure, overall amplitude was equivalent for all stimuli in the intensity staircase (e.g., at the extreme sound locations, the SPL of the dominant loudspeaker was twice that of either of the two loudspeakers when the sound was presented centrally).[4]

Procedure. Participants made left–right discriminations regarding the perceived location of sounds using two foot pedals located under their toes. Participants had to lift their right toes from the right pedal for right sounds and their left toes from the left pedal for left sounds. An orange LED placed below the fixation LED was illuminated when either pedal was lifted to indicate that a response had been registered.

The experiment consisted of two sessions of four blocks each. In each block, two staircases were completed, one beginning on the left (the left staircase), and the other on the right (the right staircase). The staircase (left or right) was chosen at random on each trial to avoid the possibility that participants might guess the location of the sound on the basis of their response to the previous trial. Each staircase began with the sound at the maximum displacement from the center. In the left staircase, left responses resulted in a displacement of the sound toward the right in the next trial belonging to that staircase (and vice versa for the right staircase). When a "right" response was made to the left staircase (or a "left" response to the right staircase), the next trial would present a sound one step back in that staircase (i.e., toward the left in the left staircase or toward the right in the right staircase); otherwise, no feedback was given after the trial. Left responses to the right staircase or right responses to the left staircase will be termed "incongruent responses." If an incongruent response was made at the extreme point in one staircase, the next trial in that staircase remained at the same point (this occurred very rarely). Note that response congruency is defined in terms of the staircase to which it has been made, and therefore a response that is objectively correct can be incongruent (i.e., responding right to the left staircase when the midpoint has been crossed). The initial step size in the staircases was two difference units, reducing to one unit within eight difference units of the center (as in Bertelson & Aschersleben's, 1998, study). Each staircase was stopped after 10 incongruent responses had been recorded, and the block finished when both staircases had been stopped.

In half of the blocks (corresponding to the bimodal condition), vibrations were delivered to the fingers in synchrony with the sound bursts. The vibrators were driven by a 200-Hz sine wave produced by a sound wave generator and amplified to a suprathreshold level. In the remainder of the blocks (corresponding to the unimodal condition), no vibrations were presented (N.B.: posture remained constant across both block types). Participants were instructed to fixate on the central LED throughout the experiment and to concentrate on the auditory task while ignoring the vibrations as much as possible during the bimodal blocks.

Design. Combining "localization cue" (time vs. intensity difference) and "modality" (bimodal vs. unimodal) resulted in four possible block types. During each of the two sessions, participants were tested once on each block type. The first two blocks were always of the same localization cue, and the last two blocks were of the other cue type. The unimodal and bimodal conditions were alternated on a block-by-block basis. There were four possible orders for the blocks in any given session. Each participant's second session began with the alternative localization cue and the alternative condition from the initial block of his/her first session.

The locations of the two loudspeaker cones were reversed between sessions to compensate for any possible acoustic difference between them. The relative positions of the loudspeakers on the first block were also counterbalanced across participants. At the beginning of the first session, a short period of training was given, allowing the participants to familiarize themselves with the task. The training consisted of two staircases with intensity difference used as the localization cue, without concurrent tactile stimuli (i.e., unimodal presentation). Each staircase was stopped after two incongruent responses in the training phase.

Results

We assessed the mean difference between the locations (ranging from the left end to the right end in 33 steps, from -16 to 16) at which incongruent responses occurred in the left and right staircases for each condition (Figure 2). This difference score reflects the range of spatial locations at which sounds can no longer be judged accurately. When time difference was used as the localization cue, there were significant differences between the unimodal and bimodal conditions [$M = 1.2$ units; $SE = .7$ in the unimodal condition vs. $M = 3.4$ units, $SE = 1.0$ in the bimodal condition, $t(7) = -3.5$, $p = .01$, two-tailed]. When intensity differences were used as the auditory localization cue, there was no effect of modality [$M = -1.1$ units, $SE = .1$ for the unimodal condition vs. $M = -0.8$ units, $SE = .4$ for the bimodal condition, $t(7) = -0.63$, $p = .55$, two-tailed].

In order to analyze further the tactile capture phenomenon observed in Experiment 1, we broke down the average location of incongruent responses by serial position for each staircase (Figure 3). This revealed several notable aspects of the serial position data. First, overall performance improved over the duration of the block in the time-difference condition, with the first incongruent responses in each staircase occurring far from each other and subsequent incongruent responses converging toward a central value (Figure 3a). This practice effect was not observed in

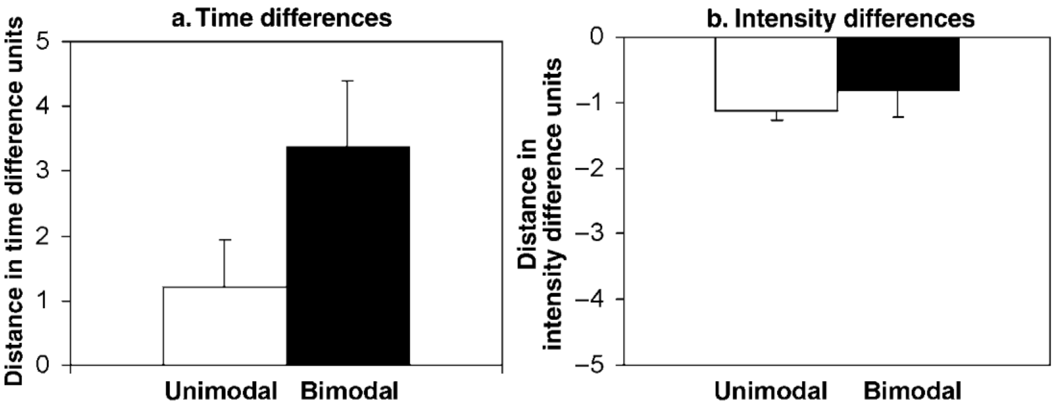

Figure 2. Mean distance between incongruent responses on the right and left staircases for sounds whose apparent location was determined by either time (a) or intensity (b) differences in Experiment 1. The results are presented separately for the unimodal and bimodal conditions.

the intensity-difference staircase (Figure 3b), where performance was very accurate from the start. Second, the differences between the bimodal and the unimodal conditions in the time-difference blocks accumulated during the initial part of the block, where performance was not at ceiling. Finally, there was a slight tendency for the staircases to converge somewhat to the left of the theoretical center (within 1 to 1.5 staircase units), although the grand average location of the last three incongruent responses was not significantly different from zero with either type of stereophonic cue [$M = -3.4$, $SD = 5.8$, $t(7) = 1.6$, $p = .140$ for the time-difference blocks, and $M = -2.1$, $SD = 3.5$, $t(7) = -1.7$, $p = .131$ for the intensity-difference blocks].

Discussion

The main result to emerge from Experiment 1 is that synchronized tactile stimulation modulated the perceived location of sounds whose position was manipulated using interaural time differences. Incongruent responses occurred further from the center when synchronized vibrations were presented at the fingertips than when no vibrations were delivered. In line with previous ventriloquism studies (e.g., Bertelson, 1998), we interpret these results to reflect an attraction of the perceived sound source toward the position of tactile stimulation (i.e., a tactile capture of audition). Note that with the present method, irrelevant tactile stimuli are neutral (located cen-

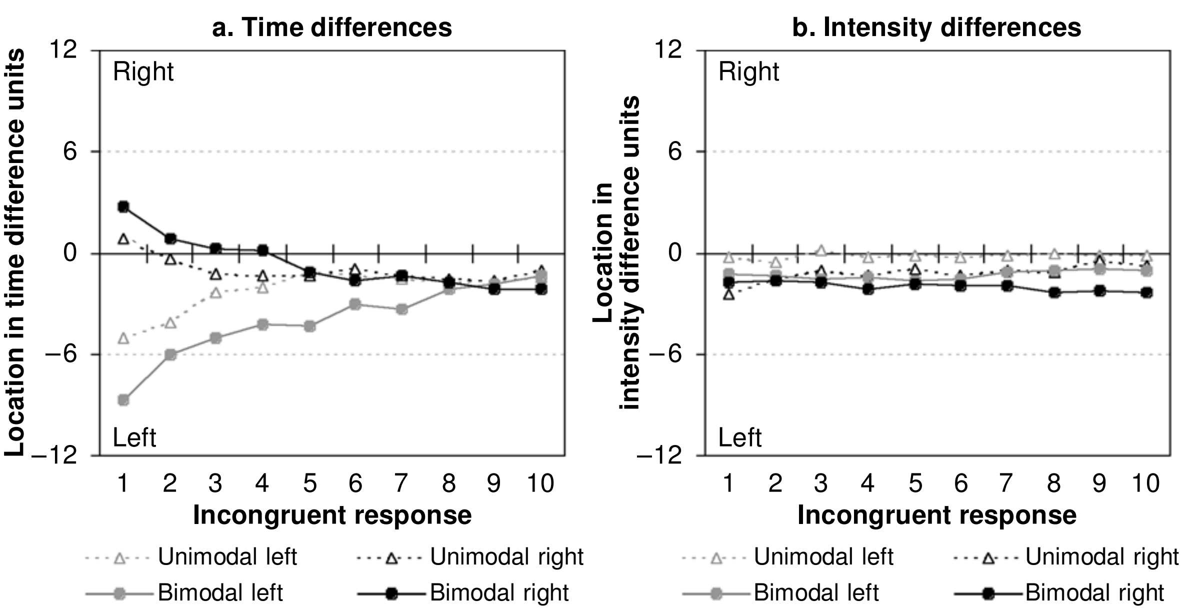

Figure 3. Mean locations of successive incongruent responses for sounds whose apparent location was determined by either time (a) or intensity (b) differences, as a function of the staircase (left or right) and the condition (unimodal or bimodal) in Experiment 1. Positive values (y-axis) refer to sounds coming from the right.

trally) with respect to the two potential responses to the target modality (left or right). Therefore, biases in the sound location task as a function of response codes activated by the vibrations are simply not possible. Instead, this result is more parsimoniously accounted for by multisensory integration of auditory and tactile information occurring in the bimodal condition, whereby the perceived location of the sounds shifted toward the location of tactile stimulation.

However, no such tactile capture of audition was reported in the blocks where sound position was manipulated via interaural intensity cues. Note that, overall, results indicate that participants found it easier to localize sounds in the intensity-difference blocks than in the time-difference blocks, a finding supported by participants' subjective reports. The intensity-difference units may have corresponded to larger angles than the time-difference units (for the values used in this experiment) and therefore allowed participants to perform at ceiling (within less than one step from the center, corresponding to 2.5º–3º). The observation that the ventriloquism effect does not occur when the sounds are easily localized has been reported in previous studies of audiovisual ventriloquism and is therefore not surprising in the present audiotactile context (e.g., Fisher, 1962; Ghahramani, 1995; Spence & Driver, 2000). Additionally, the null effect in the intensity-difference condition provides further evidence that the significant effect found in the time-difference condition was due to a genuine multisensory interaction between audition and touch, rather than to some artifact produced by our experimental paradigm.

Inspection of the time course data (Figure 3) also reveals a slight (nonsignificant) trend for responses to converge to the left of the midline, particularly in the bimodal condition. This means that participants tended to perceive the sounds as centered marginally to the left of the staircase midpoint (theoretical central position in space). This slight discrepancy (an offset of about 2.5º–5º) between the theoretical midpoint of the staircase and the perceived center is nevertheless orthogonal to our current considerations, since it affects every condition equally. Indeed, one quite remarkable fact is that an equivalent deviation between the perceived and theoretical midpoint was observed for each of the stereophonic cues. This deviation might be due to subtle asymmetries in the experimental display or in the room itself. The other interesting fact to emerge concerning the time course data is that the effect of modality condition (unimodal vs. bimodal) seems to have been due mainly to a difference in the position of the earlier incongruent responses. That is, the tactile capture of sound appears have to been greater earlier in the block. There are several possible explanations for this result.

First, it may be that audiotactile ventriloquism declined with time on the task because participants were better able to localize the sound when it was presented repeatedly. As highlighted earlier, larger ventriloquism effects occur when the localization of the sound is ambiguous or difficult (e.g., Fisher, 1962; Ghahramani, 1995; Spence & Driver, 2000). Given that overall performance improved during the course of a trial (because of practice, because of symmetrical staircases acting as reference points for one another, or because of some unknown combination of these factors), ambiguity of the sound location would decrease. Ventriloquism would then diminish during the course of a block and eventually disappear simply because sound localization becomes less ambiguous. Using a very similar procedure, Bertelson and Aschersleben (1998) observed the same pattern of data in their study of the visual capture of audition (e.g., see their Figures 1 and 2).

A second possible explanation for why the difference between bimodal and unimodal conditions occurred earlier in the block is related to a possible habituation to the vibrotactile stimuli. In particular, vibrations might have lost their relevance or saliency as participants were presented with them repeatedly. If this was the case, we might have underestimated the influence of touch on the localization of sounds, since vibrations may cease to attract the sounds once participants have habituated to them. We addressed this possibility directly in Experiment 2.

Finally, the present results could also be explained in terms of the vibrations having a distracting effect on the auditory localization task in the bimodal condition (but not in the unimodal condition, where no vibrations were present). According to this account, the distracting power of the vibrations may have been less as the trial advanced, since participants would become more accustomed to them and so perhaps be better able to ignore them. This third alternative is addressed in Experiment 3.

EXPERIMENT 2

In Experiment 2, we attempted to reduce any potential habituation to the vibrations by making them unpredictable. Vibrations were now presented randomly on only half of the trials, without participants having any advance knowledge of which trials would contain vibrotactile stimulation. This manipulation should enhance the saliency of the vibrations when they occur, and therefore make any habituation less likely. In the present experiment, vibrations were also included in the training phase to rule out an explanation of our results in terms of participants being unfamiliar with the vibrotactile stimulation, leading to a lack of concentration to the location of the sound on the first trials, where vibrations were presented. In addition, we now used only time differences to manipulate perceived sound location, given that the intensity-difference procedure seemingly rendered sound location so unambiguous as not to allow for any significant audiotactile ventriloquism. We hypothesized that if habituation accounted for the decline of tactile capture of sounds over time reported in Experiment 1, then this decline should be less marked in the present experiment (thus resulting in a capture effect that was extended in time and/or in magnitude).

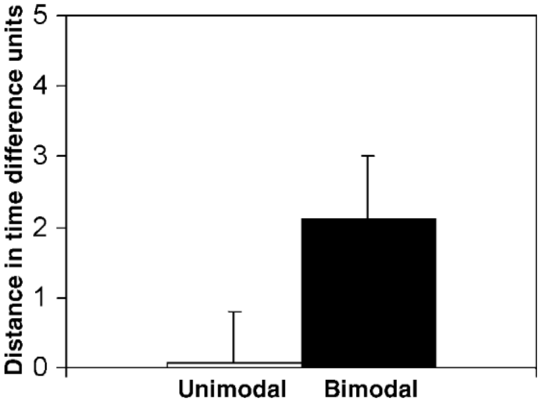

Figure 4. Mean distances between incongruent responses on the right and left staircases, separated by condition (i.e., unimodal vs. bimodal), in Experiment 2.

Method

Participants. Eight paid participants (6 female), all right-handed except one by self-report, participated in two sessions of about 20 min each. The mean age was 22 ($SD = 9$) years. All participants reported normal hearing and touch and were naive as to the purpose of the experiment.

Design. The apparatus, materials, and procedure were as in the time-difference blocks of Experiment 1, with the exception that instead of having the bimodal and unimodal conditions (i.e., with or without tactile stimulation) blocked, the occurrence of tactile stimulation was now made unpredictable by randomizing the presentation of trials from the two conditions within each block (thus, four concurrent staircases were run per block of trials). Two staircases started from the furthest position on each side (i.e., far left and far right). For each side, one staircase corresponded to the unimodal condition (without vibrations) and the other to the bimodal condition (with vibrations). The staircase presented on each trial was chosen at random from among the four possible staircases to make the presentation of vibrations on any trial unpredictable.

This experiment consisted of two sessions, each involving two blocks containing four staircases. As in Experiment 1, the locations of the loudspeakers were also switched between sessions, and the starting locations of the loudspeakers were counterbalanced across participants. The relative positions of the vibrators were also interchanged between participants. At the beginning of the first session, a training phase (two staircases, one starting from the left and one from the right) was run. Staircases in the training phase included vibrations, presented randomly on half of the trials. Each staircase in the training phase was stopped after two incongruent responses.

Results

As in Experiment 1, the average differences across participants between the locations of incongruent responses for each staircase were assessed for both conditions (Figure 4). Again, the difference between unimodal and bimodal conditions was significant [$M = 0.1$ units, $SE = .7$ in the unimodal condition vs. $M = 2.1$ units, $SE = .9$ in the bimodal condition, $t(7) = -5.0$, $p = .002$, two-tailed].

In order to test for the effects of the predictability of touch, we conducted a two-way analysis of variance (ANOVA) including the data from the time-difference blocks of Experiment 1. Predictability was included as a between-participants factor (the time difference blocks of Experiment 1 containing predictable vibrations vs. Experiment 2 containing vibrations randomly presented on 50% of the trials) and modality condition (unimodal vs. bimodal) as a within-participants variable. A significant effect of modality condition was found [$F(1,14) = 32.21$, $p < .001$], but there was no effect of the predictability of touch nor interaction between predictability and modality. The main effect of modality was still significant when we included just the data from the first two blocks of Experiment 1 alone [i.e., ensuring same number of measures per participant across experiments; $F(1,14) = 22.85$, $p < .001$].

We assessed the average location of incongruent responses at each serial position (from the 1st to the 10th) across participants and blocks (Figure 5). As Figure 5 clearly shows, the capture effect (the difference between unimodal and bimodal conditions) was still concentrated in the initial part of the block and did not extend over a longer period than in Experiment 1.

Discussion

The results of Experiment 2 replicate the main finding of Experiment 1, showing once again that sound localization can be affected by the presence of synchronized vibrations to the fingertips. In addition, the results of Experiment 2 indicate that making the vibrations unpredictable did not affect the magnitude of this tactile capture of touch. Therefore, it does not appear that simple habituation to the vibrations can explain the decrease in the capture effect as

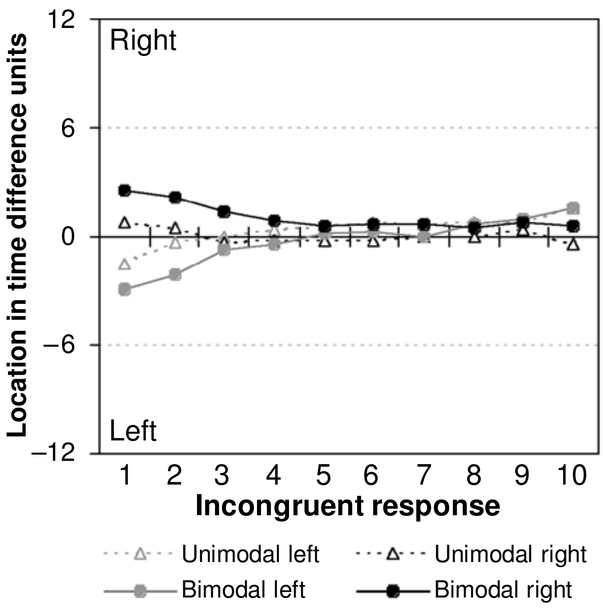

Figure 5. Mean locations of successive incongruent responses, by staircase (left or right) and condition (unimodal vs. bimodal), in Experiment 2. Positive values (*y*-axis) refer to sounds coming from the right.

the blocks advanced. It seems more likely that, given that our participants' ability to localize auditory targets improved during the block with repeated presentation of the stimuli, the ambiguity of sound location, and hence ventriloquism, declined. This suggestion fits with the hypothesis that tactile capture of audition occurs when the localization of the auditory targets is difficult or ambiguous but not when it is unambiguous, just as shown previously for audiovisual ventriloquism effects (e.g., Fisher, 1962; Ghahramani, 1995; Spence & Driver, 2000).

According to the results of Experiments 1 and 2, it seems likely that the present results reflect the fact that sounds are mislocalized in the bimodal condition because of the spatial integration of auditory and tactile events. However, as pointed out in the discussion of Experiment 1, there is an alternative account for these results, which is that they may merely reflect a generalized impairment in the performance of the sound localization task caused by the distracting effect of the tactile stimulation on the bimodal trials. The fact that Experiment 2 showed no influence of the predictability of vibrations on the magnitude of the tactile capture effect argues against this alternative, given that since Experiment 2 induced less habituation, the distraction effect should have been bigger rather than equal to that in Experiment 1. Nevertheless, one could argue that the test in Experiment 2 was not strong enough because, for instance, vibrations could have been made even more infrequent (i.e., by appearing on far fewer than half of the trials).

Therefore, it is still possible that according to the distraction hypothesis, participants may have found it harder to concentrate on the auditory task when vibrations were presented, thus leading to earlier incorrect responses in the bimodal staircases.[5] More specifically, we thought it possible that the tactile stimuli did not ventriloquize sounds toward the center, but instead simply initiated an exogenous shift of attention away from audition and toward the tactile modality. Any such shift of attention away from audition would also have been expected to reduce the accuracy of auditory discrimination responses (see Spence & Driver, 1997; Spence, Nicholls, & Driver, 2001). This alternative explanation is addressed in Experiment 3.

EXPERIMENT 3

The aim of Experiment 3 was to investigate whether the effects observed in the previous two experiments really reflect audiotactile spatial integration or instead reflect a distraction effect produced by vibrations occurring during the trial. To this end, we compared performance in the unimodal condition (no vibrations) with performance in two types of bimodal conditions—one in which vibrations occurred synchronously with every sound (as in Experiments 1 and 2) and another in which the sounds and vibrations occurred asynchronously.

There is now evidence that synchronization in the presentation of the stimuli is critical for ventriloquism to occur (e.g., Radeau & Bertelson, 1987; Slutsky & Recanzone, 2001). For example, Radeau and Bertelson (1987) showed that synchronization of inputs modulated the magnitude of audiovisual bias. In Radeau and Bertelson's study, the target and competing signals (one visual and the other auditory) were presented in any one of three different temporal configurations: continuously on, flickering at a slow tempo, or flickering at a fast tempo. The visual bias of audition and auditory bias of vision were obtained in all nine possible temporal combinations of inputs, with larger biases occurring when the two signals were presented synchronously than when they were presented asynchronously. Therefore, if the tactile capture effects observed in Experiments 1 and 2 are ventriloquism-like, desynchronizing the auditory and tactile inputs should reduce or suppress our effect. On the other hand, if the results of Experiments 1 and 2 reflect attentional distraction instead, the spatial bias should persist regardless of the desynchronization of the auditory and tactile inputs, because vibrations are still present while the sound localization task is being performed.

Method

Participants. Sixteen paid participants took part in two experimental sessions of about 20 min, each separated by a short break. All reported normal hearing and touch, and were naive as to the purpose of the experiment. The participants were separated into two groups. One group (Experiment 3a) was composed of 8 female students, 5 right-handed and 3 left-handed according to self-report, with a mean age of 19 ($SD = 2$) years. The other group (Experiment 3b) was composed of 8 students (7 females), all right-handed according to self-report and with a mean age of 19 ($SD = 0.5$) years.

Design. This experiment used the same procedure as Experiment 2, except that in the bimodal condition of Experiment 3b tactile stimulation was no longer synchronized with the sounds. A constant 61-dB (A) SPL white noise was presented from a loudspeaker cone located centrally just below the vibrators to mask any residual noise produced by their functioning. In Experiment 3a, vibrations were synchronized with the sounds as in Experiment 2. Therefore, Experiment 3a was the same as Experiment 2 but with the addition of constant white noise. In Experiment 3b, a 15-msec vibration was presented at a stimulus onset asynchrony (SOA) of −300, −200, +200, or +300 msec with respect to each of the sound bursts. The SOAs for each of the three sound–touch pairs in a trial were chosen randomly (among the four possible SOA values) and were independent of each other, thus making the vibrations asynchronous with the sounds in an unpredictable pattern.

The design was the same as in Experiment 2, with two sessions composed of two blocks of four staircases. In Experiments 3a and 3b, a training phase similar to that used in Experiment 2 was presented at the start of the first session; for participants in Experiment 3b, the vibrations were not synchronized with the sound bursts. In both cases, the training session was run with white noise (as in the experimental session).

Results

The average distance between incongruent responses on the left and right staircases was assessed for each condition (Figure 6). In Experiment 3a, the distance between staircases in the bimodal condition was significantly larger than in the unimodal condition [$M = 7.9$ units, $SE = 2.6$ in the unimodal condition vs. $M = 10.3$ units, $SE = 2.9$ in the bimodal condition, $t(7) = -3.4, p = .01$, two-tailed],

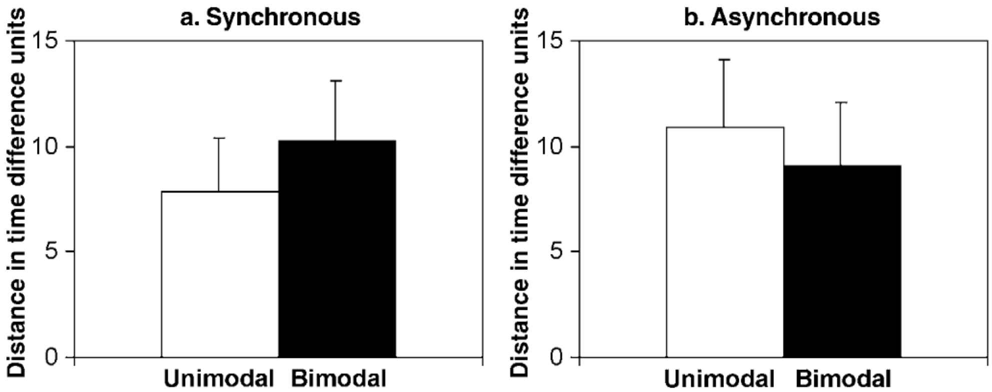

Figure 6. Mean distances between incongruent responses on the right and left staircases, by condition (unimodal vs. bimodal), in Experiments 3a (a) and 3b (b).

thus replicating the pattern of results reported in Experiments 1 and 2. No significant differences were observed between the bimodal and unimodal conditions in Experiment 3b, although a numerical trend in the reverse direction approached significance [$M = 10.9$ units, $SE = 3.2$ in the unimodal condition and $M = 9.1$ units, $SE = 3.0$ in the bimodal condition, $t(7) = 2.2$, $p = .06$, two-tailed].

An ANOVA on the mean distances between incongruent responses on the left and right staircases was conducted, including modality (unimodal vs. bimodal) as a within-participants factor and synchrony (Experiment 3a [synchronous] vs. Experiment 3b [asynchronous]) as a between-participants factor. In support of the multisensory integration account of our results, a significant interaction between modality and synchrony was found [$F(1,14) = 15.18$, $p = .002$]. This confirms that concurrent synchronized vibrotactile stimuli affected auditory localization more than did desynchronized vibrotactile stimulation. No other effects reached significance in this analysis.

In view of the data regarding the average location of incongruent responses at each serial position (Figure 7), some differences with respect to the previous experiments should be noted: (1) Overall accuracy was worse in Experiment 3 than in Experiments 1 and 2; (2) the improvement in auditory localization performance over time was modest and occurred only in the initial part of the block in Experiment 3, as compared with Experiments 1 and 2, where the improvement occurred throughout the block and the final in-

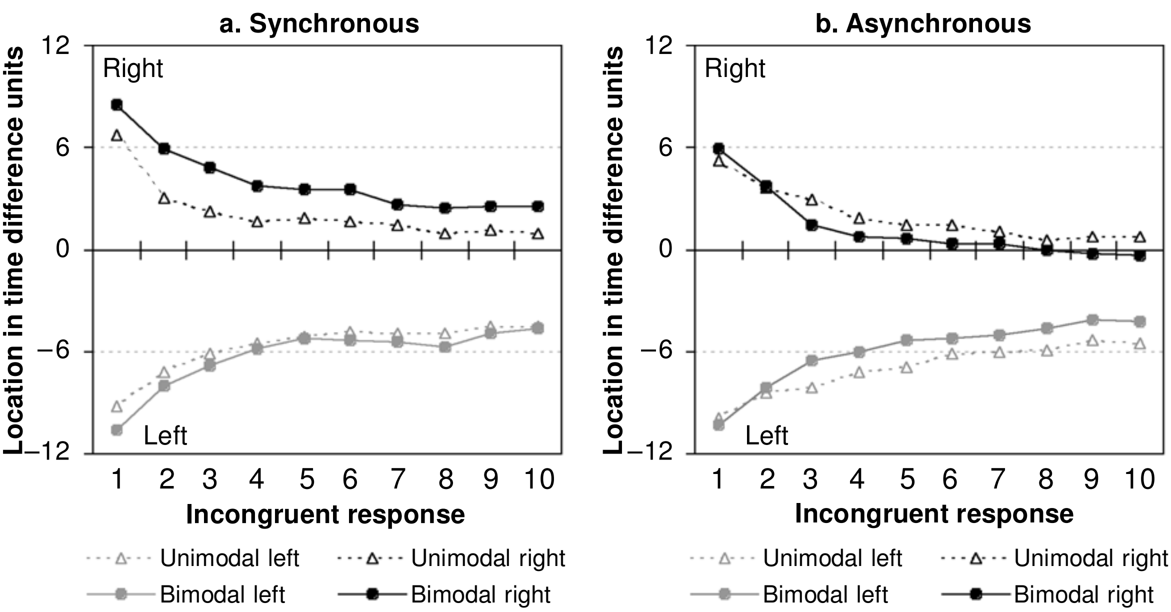

Figure 7. Mean locations of successive incongruent responses, by staircase (left or right) and condition (unimodal or bimodal), in Experiments 3a (a) and 3b (b). Positive values (*y*-axis) refer to sounds coming from the right.

congruent responses for the left and right staircase occurred very close to each other, and (3) the magnitude of the tactile capture stayed fairly constant throughout the block.

Discussion

The most important result to emerge from Experiment 3 is that only synchronized vibrations impaired sound localization performance as compared with the performance in the unimodal condition. Desynchronized vibrations had no significant effect on performance. These results support the perceptual interpretation of the tactile capture of audition given that the potential distraction from the vibrations should have been the same in both the synchronous and asynchronous conditions. This result allows us to rule out the alternative distraction account of the results from the bimodal conditions in Experiments 1 and 2.

There was a nonsignificant trend for performance in the bimodal condition to be higher than in the unimodal condition in the asynchronous staircases of Experiment 3b. This result could be explained if the asynchronous vibrations, contrary to attracting the perceived location of the sounds, provided a cue (i.e., a spatial reference point) with which to localize the center of the experimental setup. For example, Dufour, Després, Pebayle, and Brochard (2000; see also Platt & Warren, 1972) recently reported that auditory localization can be facilitated by the presentation of a central visual or auditory reference point. However, the argument that the asynchronous vibrations may have acted as a tactile reference point in our study seems unlikely given that participants already had a visual fixation point, which should have provided a more accurate (and continuously presented) reference point to indicate the center of the display. Moreover, Platt and Warren reported that proprioceptive information—provided by positioning the forefinger at the reference point—had no beneficial effect on performance in their study of auditory localization.

Alternatively, the marginal improvement in performance seen in the asynchronous audiotactile condition of Experiment 3 may have been caused by the vibrotactile stimuli acting as generalized warning, or alerting, signals. Many previous studies have shown that responses to an event in one modality are affected by the presentation of an irrelevant stimulus in another modality (e.g., Posner, 1978). For example, Posner, Nissen, and Klein (1976) have shown that discrimination responses to visual targets can be facilitated by the presentation of an auditory event at approximately the same time. Such nonspatial warning effects have been attributed to a generalized alerting effect caused by the accessory stimulus (the auditory event in Posner et al.'s study), which may serve to prepare the participant to respond to the imperative stimulus. It is possible that the desynchronized vibrotactile distractors presented in our experiments may also have acted as alerting stimuli, hence improving our participants' ability to respond to the auditory targets relative to performance in the unimodal condition.

There are several additional observations worthy of comment in Experiment 3. First, overall accuracy was lower in the present experiment than in Experiments 1 and 2, presumably due to the inclusion of white noise during the experimental sessions of the present experiment (e.g., Zwislocki, 1978). Second, little improvement over time was observed in this experiment (unlike in Experiments 1 and 2), which may again reflect the increased difficulty of sound localization caused by the presentation of background noise (perhaps rendering any practice effects less prominent). Third, the capture effects observed in Experiment 3a varied very little over time compared with the remarkable reduction in capture observed in the time course of Experiments 1 and 2. This fact, when taken together with the reduced effects of practice in this experiment, again supports the idea that only when sound localization is difficult or ambiguous are capture effects observed. In the present experiment, sound location remained ambiguous throughout the whole block, and accordingly the ventriloquism illusion persisted throughout.

Experiment 4 was designed to characterize further the phenomenon of capture reported here and to examine more precisely whether or not the magnitude of capture effects would be modulated by the voluntary division of attention between the two modalities compared with the situation in which solely the auditory modality is attended to (i.e., focused attention), as in the previous experiments.

EXPERIMENT 4

In order to assess whether the direction of voluntary attention modulates the tactile capture of audition, we introduced two different task conditions in Experiment 4. In one condition, participants were required to respond to the location of the sounds on every trial (i.e., participants were instructed to focus their attention solely on audition, as in the previous experiments). In the other condition, participants were required to respond to the sound location but also to detect occasional tactile targets (i.e., participants were required to divide their attention between audition and touch; see Lloyd, Merat, McGlone, & Spence, 2000). If the tactile capture of audition is modulated by voluntary (or endogenous) attention, then it would be predicted that the mislocalization of sounds toward the vibrations should become more prominent as attention is partially directed toward touch (and away from audition).

Method

Participants. Eleven volunteer or paid participants (8 female), by self-report all right-handed except 1, participated in this 1-h experiment. The mean age was 26 (SD = 3) years. All participants reported normal hearing and touch, and were naive as to the purpose of the experiment.

Procedure and Design. The apparatus and basic procedures were as in Experiment 2: Unimodal and bimodal trials were randomly intermingled, which was again achieved by the concurrent running of four staircases, and the experiment was conducted without background white noise. In order to shorten the experiment, each of the four staircases was stopped after 5 incongruent responses (instead of 10 as in the previous experiments). Two types of block were presented. One was exactly as reported in Experiment 2; attention was directed solely to the auditory modality by the instructions given (focused attention block). In the other block type (divided-attention block), the two bimodal staircases contained an additional 10% of randomly distributed *odd* trials containing three sound bursts but only two vibrations. The remaining 90% of trials were composed of three sound bursts and three synchronous vibrations. Participants were instructed to localize the sound in the "standard" trials (with either three vibrations or no vibrations) and to respond to the "odd" trials by lifting both pedals at the same time. Therefore, in this group, participants had to attend to the vibrations as well as to the sounds in order to comply with the instructions (note that Bertelson & Aschersleben, 1998, always used catch trials in the visual modality in their study). In the divided-attention block, the odd trials were not taken into account in assessing the location of the sound on the next trial in that staircase. Participants were tested three times on each of the two block types. Attention conditions (focused auditory attention vs. divided attention) were alternated on a block-by-block basis, and the order of presentation was counterbalanced across participants.

Two short training phases were given at the beginning of the experiment. Both involved two staircases of randomly intermingled unimodal and bimodal trials, stopped after two incongruent responses. The first training block was always of the auditory-attention-alone condition type and the second of the divided-attention type (for the training, the proportion of "odd" trials was increased to 30% of the bimodal trials). The relative position of the loudspeakers and vibrators was counterbalanced across participants. A first set of 8 participants was run, and data from participants who failed to report more than 20% of the odd trials in the divided-attention condition were discarded. Three participants did not reach this response accuracy criterion and were replaced by 3 additional participants who all reached the criterion.

Results

The average distance between incongruent responses on the left and right staircases were assessed for each condition (Figure 8). The results were analyzed in a two-way ANOVA with modality (unimodal vs. bimodal) and attention (focused vs. divided attention) as within-participants factors. As in our previous experiments, there was a significant main effect of modality [$F(1,7) = 14.23, p = .007$]. Critically, neither the effect of attention ($F < 1$) nor the interaction between modality and attention [$F(1,7) = 1.55, p = .25$] approached significance.

The mean position of the incongruent responses across participants and blocks was also assessed for both attention conditions in this experiment (Figure 9). As in Experiments 1 and 2, in which no white noise was used, the serial position data reveals that there was a practice effect and that the capture effect correlated negatively with this temporal improvement. Note, however, that unlike the results of Experiments 1 and 2, the present results indicated that the convergence of the staircases was not complete. This is a natural consequence of the shortening in the number of incongruent responses allowed in this experiment, by which the overall number of trials was decreased, and hence the chances of reaching the midpoint of the staircase within the course of one block were also reduced.

Discussion

The results of Experiment 4 showed an alteration of sound localization performance when synchronous vibrations were presented at the centrally placed fingertips, thus replicating the effect reported previously in Experiments 1, 2, and 3a. Importantly, this capture effect was unaffected by whether participants attended solely to audition or divided their attention between audition and touch. The finding that the allocation of attention between modalities has little effect on the tactile capture of audition is inconsistent with certain older studies of the effects of attention on ventriloquism (e.g., Canon, 1970, 1971; Radeau & Bertelson, 1976), but our findings coincide more closely with recent data showing that immediate audiovisual intersensory biases are independent of the direction of either voluntary (Bertelson, Vroomen, de Gelder, & Driver, 2000) or reflexive (Vroomen, Bertelson, & de Gelder, 2001) visual attention.

It should be noted, however, that one could argue that the tactile task in the present experiment might have been too easy and consequently demanded too few attentional resources to produce a significant effect of attention. Two lines of evidence argue against this possibility. First, the fact that 3 out of 11 participants performed poorly on the

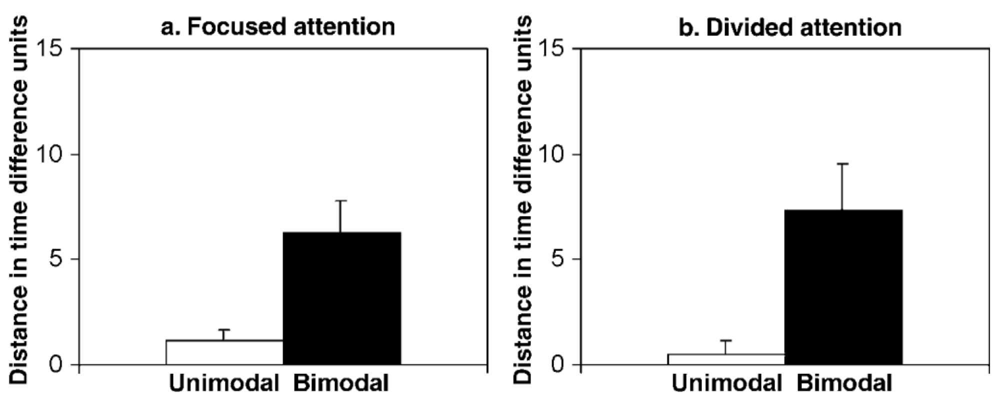

Figure 8. Mean distances between incongruent responses on the left and right staircases, in the focused attention blocks (a) and in the divided-attention blocks (b), presented by condition (unimodal vs. bimodal), in Experiment 4.

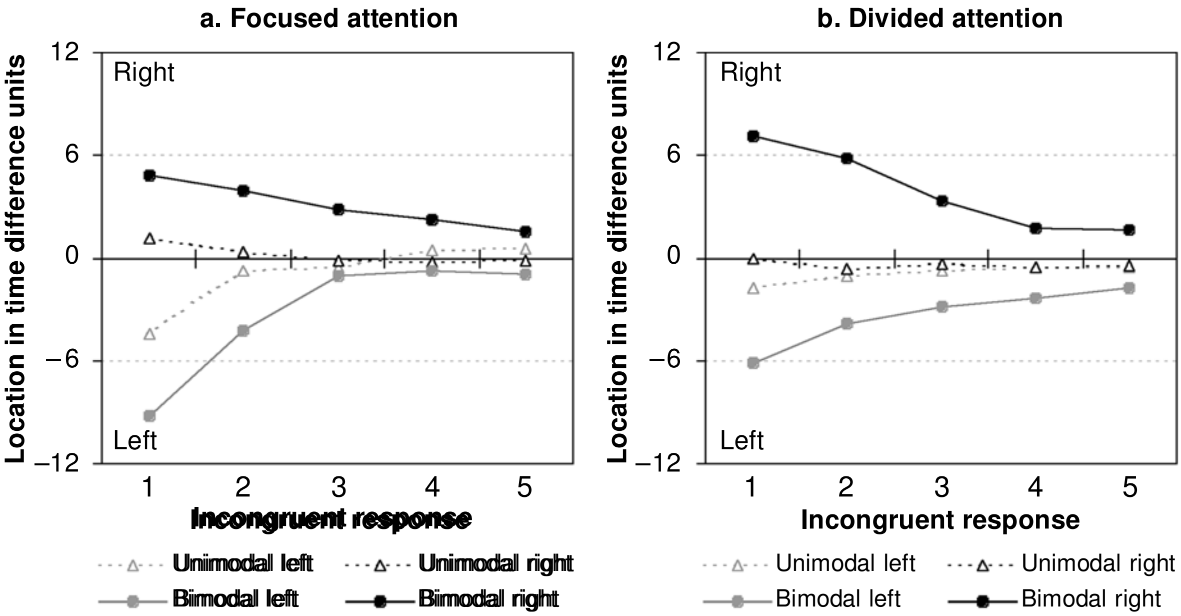

Figure 9. Mean locations of successive incongruent responses in the focused-attention blocks (a) and in the divided-attention blocks (b) in Experiment 4. Results are separated by staircase (left or right) and condition (unimodal vs. bimodal). Positive values (*y*-axis) refer to sounds coming from the right.

tactile task (i.e., missed more than 20% of the trials with only two vibrations) suggests that the tactile task could not be performed easily. Second, the fact that all participants included in the analyses performed at over 80% correct on the tactile task indicates that they were not simply ignoring the tactile modality. However, the conclusion that attention does not affect audiotactile integration must remain tentative at this stage. Ideally, future studies should provide an independent measure of the distribution of attention across the different sensory modalities to ensure that instructions to differentially distribute attention between the various sensory modalities, and/or the introduction of a secondary task, result in a demonstrable redistribution of the participants' attention between the modalities (see Spence, Shore, & Klein, 2001, on this point).

GENERAL DISCUSSION

Experiments 1–4 demonstrate that the perceived location of a sound can be affected by the simultaneous presentation of tactile stimulation. In these experiments, incongruent responses (e.g., answering "left" to a sound belonging to the right staircase) occurred further away from the midpoint when vibrations were presented at the centrally located fingertips than with desynchronized or no vibrations. These data indicate that there is a spatial interaction between auditory and tactile localization such that concurrent tactile stimulation biases the perceived location of sound toward that of touch. Experiment 3 allowed us to rule out the hypothesis that this performance impairment might be due to a simple distracting effect of the vibrations, because desynchronizing the auditory and tactile stimuli abolished the capture effect. Experiment 4 provides preliminary evidence that tactile capture of audition cannot simply be modulated by dividing attention between the two modalities as opposed to being focused just on the auditory modality, suggesting no involvement of endogenous (or voluntary) attention on multisensory integration—at least for the relatively sparse stimulus arrays used in the present study. Note also that in Experiments 1 and 3 the staircases tended to converge toward the left of the theoretical midline (where the two sounds are in phase). This cannot be due to the differential strength of the two vibrators because they were switched across participants (in Experiments 2–4). The source of these marginal biases is unclear, but it is worth pointing out that the crossmodal capture effects were observed independently of their presence.

We have attributed the behavioral capture effects in the present experiments to the multisensory integration of spatially discrepant tactile and auditory stimuli, which in turn leads to the mislocalization of auditory stimuli. An alternative explanation for our results is that the presentation of the tactile distractor might simply lead to an exogenous shift of attention toward the tactile modality (i.e., a nonspatial orienting effect), and that this might explain why auditory localization judgments were impaired by the presentation of the central tactile stimulus (e.g., Harvey, 1980; Spence, Nicholls, & Driver, 2001; Turatto, Benson, Galfano, Gamberini, & Umiltà, in press). Our finding that the effect of touch on auditory localization judgments occurs only when the stimuli are presented synchronously, but not when they are presented asynchronously (although still overlapping in time) with the auditory stimuli, argues against this alternative account.

It should be noted, however, that the presentation of the tactile stimulus may also have led to an exogenous (or reflexive) shift of *spatial* attention toward the cued location (see Spence & Driver, 1997; Spence et al., 1998), which in turn may have contributed to the mislocalization of the sound toward the location of the tactile event (i.e., toward the focus of spatial attention, rather than toward the tactile event per se). Although the empirical research on crossmodal links in exogenous spatial attention and on multisensory integration describe these two phenomena independently,[6] distinguishing between the two is not trivial (see Macaluso, Frith, & Driver, 2000, 2001; McDonald, Teder-Sälejärvi, & Ward, 2001). Nevertheless, it is clear from the present study that touch can modulate the subjective localization of a sound in a manner that is similar to what is seen when vision pulls auditory localization in the ventriloquism effect.

The psychophysical staircase paradigm used in the present study provides a reliable demonstration of crossmodal interactions in spatial localization between audition and touch, just as reported previously by Bertelson and Aschersleben (1998) for the audiovisual situation. Visual inspection of Bertelson and Aschersleben's results suggests that vision biases audition more than touch does, at least in the case of flashing lights or brief tactile stimulation synchronized with sounds bursts. However, the bigger effect on auditory localization obtained with a flashing light in their study, as compared with brief tactile stimulation in the present study, could be accounted for by slight differences between the two studies that make any direct comparison impossible. For example, Bertelson and Aschersleben used the location of response reversals as an indicator of performance, whereas we used the location of incongruent responses. Perhaps more importantly, there are also possible differences in the relative intensity of the stimuli used; that is, the visual attractor may have been more intense than the tactile stimulus used here, or the sound bursts of a lower amplitude than those used here (see Radeau, 1985, for a discussion of the effect of relative stimulus intensity on ventriloquism). Therefore, although initially it might appear that the visual capture of audition is larger than tactile capture of audition, the evidence is at present inconclusive on this point.

The behavioral audiotactile interactions reported in the present study reveal that exposure to a spatial conflict between the two inputs is resolved by means of an intersensory bias in perceived stimulus location that is similar to that observed for other modality pairings. This bias seems to be strongly dependent on the synchronization of the inputs, which again resembles other modality pairings (for the audiovisual case, see Bertelson & Aschersleben, 1998; Radeau & Bertelson, 1987; Slutsky & Recanzone, 2001). The biasing of auditory location by concurrent tactile inputs also seems to occur more readily when the sounds are difficult to localize (see Ghahramani, 1995; Spence & Driver, 2000). We believe that this relationship between ambiguity in sound localization and ventriloquism explains the correlation between the increasing precision of auditory localization during the course of the block (probably as a consequence of practice with the localization task) and the amount of capture. This correlation was observed clearly in Experiments 1, 2, and 4. The fact that the magnitude of ventriloquism changed little over the course of blocks where only a very small improvement in the localization task occurred (Experiment 3a) adds support to the proposed relationship between sound localization accuracy and ventriloquism.

The question arises as to the level of processing at which the audiotactile integration demonstrated in the present study takes place. The paradigm used here was designed to reveal the genuine perceptual component of any audiotactile interaction, and so an early level of processing is likely. This interpretation is consistent with the results of a recent event-related potential (ERP) study in humans (Foxe, Morocz, Murray, Higgins, Javitt, & Schroeder, 2000) showing that audiotactile integration can occur as early as at the level of the somatosensory and auditory cortices. It is tempting to conclude that this audiotactile interaction occurs in a region of the brain (either cortical or subcortical) where a corepresentation of auditory and somatosensory space exists. Indeed, neurophysiological studies have identified multisensory audiotactile neurons in a variety of animals, both at the cortical and subcortical levels (see Foxe et al., 2000; Schroeder, Lindsley, Specht, Marcovici, Smiley, & Javitt, 2001; Stein & Meredith, 1993). However, as discussed earlier, visual maps seem to be dominant over other sensory modalities, and therefore an intermediate involvement of vision remains possible. One potential interpretation is that either secondary auditory cortical areas or parietal associative areas adjust spatial inputs from the primary auditory cortex to coincide with visual spatial inputs. Recanzone (1998) argued that frequency-selective regions of the cortex (such as the primary auditory cortex) contribute to audiovisual interactions, since ventriloquism aftereffects are not transferable across frequencies. This provides support for a cortical substrate of crossmodal spatial interactions such as ventriloquism and our tactile capture of audition that involve early representations of perceptual space.

To summarize, the main finding of the present study is that concurrent tactile stimulation modulates the perceived location of sounds. This novel result has been replicated in four experiments and has been shown to be eliminated using the same experimental paradigm simply by desynchronizing the vibrations. It has been argued that the most likely account for the present data is an interaction at a perceptual level between tactile and auditory spatial information. Additionally, our results suggest that this interaction may be independent of the allocation of voluntary attention between different sensory modalities.

REFERENCES

Auerbach, C., & Sperling, P. (1974). A common auditory-visual space: Evidence for its reality. *Perception & Psychophysics*, **16**, 129-135.

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

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

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.

Blauert, J. (1983). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, **391**, 756.

Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, **84**, 141-147.

Canon, L. K. (1971). Directed attention and maladaptive 'adaptation' to displacement of the visual field. *Journal of Experimental Psychology*, **88**, 403-408.

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.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology*, **75**, 485-491.

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

Dufour, A., Després, O., Pebayle, T., & Brochard, R. (2000). *Visual and auditory facilitation in auditory spatial localization.* Manuscript submitted for publication.

Eimer, M., Cockburn, D., Smedley, B., & Driver, J. (2001). Crossmodal links in endogenous spatial attention are mediated by common external locations: Evidence from event-related brain potentials. *Experimental Brain Research*, **139**, 398-411.

Fisher, G. H. (1962). Resolution of spatial conflict. *Bulletin of the British Psychological Society*, **46**, 3A.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 1030-1044.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, **10**, 77-83.

Freedman, S. J., & Wilson, L. (1967). Compensation for auditory rearrangement following exposure to auditory-tactile discordance. *Perceptual & Motor Skills*, **25**, 861-866.

Ghahramani, Z. (1995). *Computation and psychophysics of sensorimotor integration.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Gilkey, R. H., & Anderson, T. R. (1997). *Binaural and spatial hearing in real and virtual environments.* Mahwah, NJ: Erlbaum.

Groh, J. M., & Sparks, D. L. (1996). Saccades to somatosensory targets. III. Eye-position-dependent somatosensory activity in primate superior colliculus. *Journal of Neurophysiology*, **75**, 439-453.

Harvey, N. (1980). Non-informative effects of stimuli functioning as cues. *Quarterly Journal of Experimental Psychology*, **32**, 413-425.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation.* New York: Wiley.

Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, **309**, 345-347.

Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: Sound biased touch. *Current Biology*, **8**, R190.

Lloyd, D. M., Merat, N., McGlone, F., & Spence, C. (2000). *Crossmodal links in covert endogenous spatial attention between audition and touch.* Manuscript submitted for publication.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, **289**, 1206-1208.

Macaluso, E., Frith, C. D., & Driver, J. (2001). A reply to McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. Multisensory integration and crossmodal attention effects in the human brain. *Science*, **292**, 1791-1792.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, **407**, 906-908.

McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, **292**, 1791.

Merat, N., Spence, C., Withington, D. J., & McGlone, F. (1999). Audiotactile links in focused and divided spatial attention. *Society for Neuroscience Abstracts*, **25**, 356.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments, & Computers*, **29**, 322-327.

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological Science*, **11**, 353-359.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, **6**, 203-205.

Platt, B. B., & Warren, D. H. (1972). Auditory localization: The importance of eye movements and a textured visual environment. *Perception & Psychophysics*, **12**, 245-248.

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

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, **83**, 157-171.

Radeau, M. (1985). Signal intensity, task context, and auditory-visual interactions. *Perception*, **14**, 571-577.

Radeau, M., & Bertelson, P. (1974). The aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology*, **26**, 63-71.

Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, **20**, 227-235.

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

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, **95**, 869-875.

Rock, I., & Harris, C. S. (1967, May 17). Vision and touch. *Scientific American*, **216**, 96-104.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, **85**, 1322-1327.

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

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

Spence, C., & Driver, J. (1997). On measuring selective attention to a specific sensory modality. *Perception & Psychophysics*, **59**, 389-403.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism. *NeuroReport*, **11**, 2057-2061.

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

Spence, C., Nicholls, M. E. R., Gillespie, N., & Driver, J. (1998).

Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, **60**, 544-557.

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

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

Turatto, M., Benson, F., Galfano, G., Gamberini, L., & Umiltà, C. (in press). Nonspatial attentional shifts between audition and vision and a new look at visual dominance. *Journal of Experimental Psychology: Human Perception & Performance*.

Urbantschitsch, V. (1880). Über den Einfluss einer Sinneserregung auf die übringen Sinnesempfindungen [On the influence of one sensory percept on the other sensory percepts]. *Archiv für die Geschicte der Physiologie*, **42**, 154-182.

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

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, **76**, 1246-1266.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology*, **48**, 242-259.

Warren, D. H. (1970). Intermodality interactions in spatial localization. *Cognitive Psychology*, **1**, 114-133.

Welch, R. B. (1999). The advantages and limitations of the psychophysical staircases 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., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 25-1 to 25-36). New York: Wiley.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 45-76). Washington, DC: American Psychological Association.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 73-103). Cambridge, MA: MIT Press.

Zwislocki, J. J. (1978). Masking: Experimental and theoretical aspects of simultaneous, forward and backward masking. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol 4. Hearing* (pp. 283-336). New York: Academic Press.

NOTES

1. The term *capture* is used here to denote the fact that, in a crossmodal conflict situation, information in one modality prevails over (or substantially modulates) the perception of information in the other modality. This use of the term to describe such intersensory bias effects (as well as in other research in the intersensory bias literature; see Bertelson, 1998, for a review) has a long history in experimental psychology (e.g., Posner, Nissen, & Klein, 1976; Rock & Harris, 1967) and should not be confused with the more recent—and qualitatively different—use of this term in the attentional control settings literature to describe the contingent capture of attention by salient visual cues (see, e.g., Folk, Remington, & Johnston, 1992; Yantis, 1996, 2000).

2. One would not expect the results of either Freedman and Wilson's (1967) or Pick et al.'s (1969) study to be modulated by whether or not the auditory and tactile stimuli were simultaneous, a manipulation used by many researchers to demonstrate that their effects reflect the consequences of multisensory integration and not some other factor (see, e.g., the present Experiment 3).

3. Stereophonic cues were used in the present study to replicate the conditions used in Bertelson and Aschersleben's (1998) previous study as closely as possible. Although the quality of the sound location induced with stereophonic cues is not as high as that generated by an actual array of loudspeaker cones, it nevertheless suffices for the present purposes. First, stereophonic cues produce reliable and consistent localization performance from participants according to the results of all experiments in the present study (see Blauert, 1983; Gilkey & Anderson, 1997, for further details on induced sound localization). Second, several previous audiovisual studies have reported that the degree of ventriloquism of auditory events toward simultaneously presented visual events increases as the relative localizability of the sounds decrease (e.g., Fisher, 1962; Ghahramani, 1995; Spence & Driver, 2000). Therefore, the use of stereophony in the present study was designed to provide a better opportunity for demonstrating auditory capture than if a more localizable sound source had been used. Finally, the use of stereophony also allowed us to eliminate any idiosyncratic differences between particular loudspeaker cones as the underlying cause of any behavioral effects reported.

4. There is no known relationship between time differences and intensity differences in the free field.

5. It should be noted that the absence of any effect of the vibrations with the intensity-difference paradigm in Experiment 1 argues against the distraction hypothesis. It seems unlikely that vibrations could have distracted participants when asked to localize a sound with a time difference between the two loudspeakers, but not when localizing sounds with an intensity difference. Nevertheless it remains possible that localization judgments could be easy enough in the intensity-difference case to be unaffected by the distraction induced by the vibrations.

6. Neurophysiologists typically describe interactions between stimuli presented in different sensory modalities at approximately the same time in terms of multisensory integration (e.g., Stein & Meredith, 1993), whereas cognitive psychologists typically describe similar results in terms of crossmodal, or multisensory, links in exogenous spatial attention (e.g., Driver & Spence, 2000; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Spence & Driver, 1997; Ward, 1994).

(Manuscript received July 25, 2000;
revision accepted for publication September 15, 2001.)