

# The ventriloquist effect does not depend on the direction of automatic visual attention

JEAN VROOMEN

*Tilburg University, Tilburg, The Netherlands*

and

PAUL BERTELSON and BEATRICE DE GELDER

*Tilburg University, Tilburg, The Netherlands  
and Université Libre de Bruxelles, Brussels, Belgium*

Previously, we showed that the visual bias of auditory sound location, or ventriloquism, does not depend on the direction of deliberate, or *endogenous*, attention (Bertelson, Vroomen, de Gelder, & Driver, 2000). In the present study, a similar question concerning automatic, or *exogenous*, attention was examined. The experimental manipulation was based on the fact that exogenous visual attention can be attracted toward a *singleton*—that is, an item different on some dimension from all other items presented simultaneously. A display was used that consisted of a row of four bright squares with one square, in either the left- or the rightmost position, *smaller* than the others, serving as the singleton. In Experiment 1, subjects made dichotomous left-right judgments concerning sound bursts, whose successive locations were controlled by a psychophysical staircase procedure and which were presented in synchrony with a display with the singleton either left or right. Results showed that the apparent location of the sound was attracted *not* toward the singleton, but instead toward the big squares at the opposite end of the display. Experiment 2 was run to check that the singleton effectively attracted exogenous attention. The task was to discriminate target letters presented either on the singleton or on the opposite big square. Performance deteriorated when the target was on the big square opposite the singleton, in comparison with control trials with no singleton, thus showing that the singleton attracted attention away from the target location. In Experiment 3, localization and discrimination trials were mixed randomly so as to control for potential differences in subjects' strategies in the two preceding experiments. Results were as before, showing that the singleton attracted attention, whereas sound localization was shifted away from the singleton. Ventriloquism can thus be dissociated from exogenous visual attention and appears to reflect sensory interactions with little role for the direction of visual spatial attention.

Perceptual processing has generally been studied in one specific sense modality at a time. Yet, most events in real life produce stimulation that impinges simultaneously on several modalities. Evidence that the system somehow combines corresponding inputs from separate modalities has come mainly from studies with *conflict situations*, in which intermodal discordances are created on some dimension of the inputs, such as location or identity, while other dimensions are kept coherent. The best known case is, of course, visuoproprioceptive rearrangement, in which the normal correspondence between the felt location and the seen location of some body part is altered through an optical device (Held, 1965; Welch, 1978).

Presenting synchronous auditory and visual events in somewhat separate locations creates another type of con-

flict situation. Such audiovisual (AV) conflict is convenient for experimental analysis, because here, unlike in cases involving proprioception, both the location and the timing of the inputs can easily be controlled. Reactions to AV spatial conflict are currently designated by the term *ventriloquism*, because one of their most spectacular everyday examples is the illusion created by performing ventriloquists that the speech they produce without visible facial movements comes from a puppet they agitate in approximate synchrony with the speech.

Experimental studies of ventriloquism have dealt with on-line effects observed in immediate reaction to the conflicting inputs and also with off-line aftereffects of exposure to these inputs. The most important on-line effect is *cross-modal bias*, which occurs when a subject is instructed to localize inputs in one modality while ignoring spatially discordant inputs in the other modality. The response is then shifted in the direction of the competing data, despite instructions to focus on the target ones (e.g., Bermant & Welch, 1976; Bertelson & Radeau, 1981; Bertelson, Vroomen, de Gelder, & Driver, 2000; Pick, Warren, & Hay, 1969; Radeau & Bertelson, 1987). Most on-

---

P.B.'s participation in this work was supported by the Ministry of Scientific Research of the Belgian French-Speaking Community (Concerted Research Action 96/01-2037) and by the Belgian National Fund for Collective Fundamental Research (Contract 2.45.39.95). Correspondence concerning this article should be addressed to J. Vroomen, Department of Psychology, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, The Netherlands (e-mail: j.vroomen@kub.nl).

line studies have examined only the visual bias of auditory location, but a smaller auditory bias of visual location has been obtained when this was also considered (Bertelson & Radeau, 1981, Experiment 1, and 1987; but see Radeau, 1985, for a failure to replicate). Off-line aftereffects are demonstrated by having subjects localize unimodal target stimuli before and after some exposure to constant intermodal conflict and consist in shifts of the reported location of the targets toward the place occupied by the conflicting data during the exposure period (e.g., Canon, 1970; Radeau & Bertelson, 1974, 1976). Radeau and Bertelson obtained substantial aftereffects not only in audition, but also in vision, provided the visual attractors were presented against a homogeneous dark background. For both on-line and off-line manifestations, there is convincing evidence that their occurrence is contingent on the degree of synchronization between inputs in the two modalities (e.g., Bertelson & Aschersleben, 1998; Choe, Welch, Gilford, & Juola, 1975; Radeau & Bertelson, 1977, 1987).

The main reason to be interested in these phenomena is that they can help understand those processes through which a coherent internal representation of external space is established and maintained. Cross-modal bias may serve the ecologically useful role of attenuating the effects of variability, owing to noise or drift, in the performance of modality-specific localization processes by cross-reference to the other modality. This notion is consistent with the bigger weight in the bias phenomena of visual data, whose spatial resolution is far better than that of the auditory data. Regarding aftereffects, they have generally been seen as reflecting a process of recalibration of unimodal localization processes, by which intermodal spatial coherence is restored when such causes as body growth or spontaneous drift in one or both processes have been disturbing it.

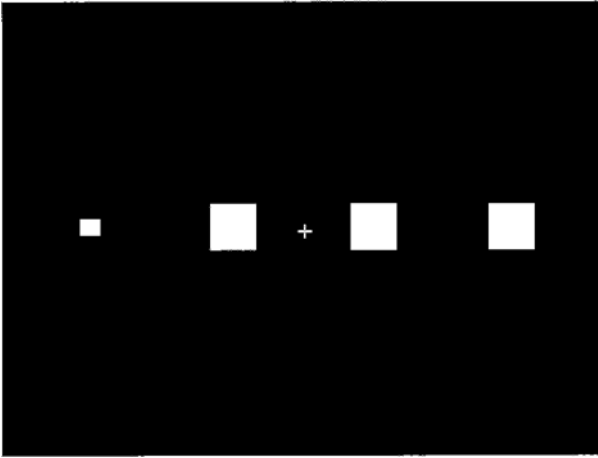
The preceding considerations involve an implicit assumption that the experimental evidence reflects the operation of genuine perceptual processes, and not just postperceptual response strategies adopted to satisfy the demands of particular laboratory tasks. The possibility of contamination by the latter kind of factor is especially strong for cross-modal bias, because the selective localization task, through which that effect has usually been demonstrated, leaves the subject free to resort to any strategy when instructed to ignore distractors of whose presence he or she is fully aware. The possibility is also increased in those studies in which the subject is given information (either true or deceitful) regarding the sources of the bimodal input (e.g., Radeau & Bertelson, 1974; Warren, Welch, & McCarthy, 1981; Welch, 1972) or those in which that input simulates familiar events known for producing correlated auditory and visual inputs, such as speech and the speaking face of the talker (Radeau & Bertelson, 1977, Experiments 2 and 3; Warren et al., 1981) or the sight of noisy objects and their typical noise (Jack & Thurlow, 1973; Jackson, 1953; Pick et al., 1969; Radeau & Bertelson, 1977, Experiment 1, 1978).

On-line effects that seem immune from postperceptual influences have recently been reported by Bertelson and Aschersleben (1998; see also Bertelson, 1999). Measuring the apparent location of sound bursts by means of psychophysical staircases, the authors were able to show that this location was attracted toward a simultaneously flashing light at a stage of the experiment at which the subject was not aware of the spatial separation between the sounds and the flashes and thus could not possibly choose a consistent response strategy based on knowledge of that separation. Results that are convergent with the preceding ones have been obtained by Driver (1996) by an indirect method, in which the visual bias of auditory location was measured in the classical "cocktail party" situation through its effect in facilitating the focusing of attention on one of two simultaneous spoken messages. Since the task made no reference to location, there was little danger that the effect was based on a voluntary localization strategy.

These data show that at least part of the ventriloquist effects obtained under less controlled conditions is due to automatic perceptual processes. They, of course, do not exclude an additional role in those conditions for factors of a more cognitive nature. This possibility is, for the time being, the object of some controversy (see Bertelson, 1999; Vroomen, 1999; Welch, 1999). On the one hand, the evidence for the effectiveness of cognitive factors is not as strong as is sometimes assumed. Effects of instructions were reported in some studies (Radeau & Bertelson, 1974; Warren et al., 1981) and not in other ones (Radeau & Bertelson, 1978), and similarly the effects of context realism were obtained in some studies (Jack & Thurlow, 1973; Jackson, 1953; Radeau & Bertelson, 1977, Experiment 3), but not in other ones (Radeau, 1992; Radeau & Bertelson, 1977, Experiments 1 and 2). On the other hand, for the positive findings, the question of the processing level at which they originated, perceptual or postperceptual, has not been examined so far, but it should be before any strong conclusions are drawn.

In discussions focused on the contrast between perceptual and more cognitive accounts of ventriloquism, one possibility, which has largely been ignored, is a role for the *direction of attention*. Spatial attention is not a purely perceptual factor, since it can be controlled deliberately, in response to informative cues to location, nor is it a wholly cognitive one, since it can also be captured automatically by uninformative salient stimuli (Egeth & Yantis, 1997; Posner & Cohen, 1984). This now well-established distinction between deliberate, or *endogenous*, and automatic, or *exogenous* control, implies that consideration of the possible effect of the direction of attention on ventriloquism can broaden the debate about underlying mechanisms beyond the perceptual versus cognitive dichotomy.

The possibility that spatial attention might play a role in ventriloquism was raised recently by results demonstrating a linkage between the mechanisms controlling the direction of attention in audition and in vision. For in-



**Figure 1.** An example of one of the displays, with the singleton square on the left.

stance, Ward (1994) observed that a visual cue displacing the focus of attention in the visual modality produced a corresponding shift in the auditory modality (Ward, 1994). Other cross-modal links, in both endogenous and exogenous attention, have been described as well (Driver & Spence, 1994; Spence & Driver, 1996, 1997). Could such intermodal attentional links be one of the determinants of ventriloquist effects, as well as of other cross-modal interactions?

In an earlier study, we examined whether ventriloquism depends on where, on the visual attractor or elsewhere, *endogenous* visual attention is focused (Bertelson et al., 2000). In one experiment, we used the selective localization task by which immediate cross-modal biases have generally been investigated (Bertelson & Radeau, 1981; Pick et al., 1969). The auditory targets, which the subject had to localize, were trains of sound bursts delivered at variable locations around the center of the display, and the visual attractors, which he or she was instructed to ignore, were bright squares flashed synchronously to the left or the right of the sounds. The experimental manipulation was to have the subject, in addition to localizing the sounds, monitor for occasional occurrences of a catch stimulus, either in the center of the display or in the lateral square. According to the attentional hypothesis, the visual bias of auditory target location should have been stronger when the attractor square was monitored than when the center was monitored. Strong biases were obtained, but contrary to the hypothesis, ventriloquism effects were of equal strength in the two attentional conditions.

In another experiment, two squares, one on either side of the middle of the display, were flashed in synchrony with the sounds, and the subject had to monitor either the left or the right square for the same catch stimulus. On some trials, the two squares had equal sizes, and no visual bias occurred in any of the two monitoring conditions. On

other trials, the squares were of unequal sizes, and there was a significant bias in the direction of the bigger square, but again regardless of where monitoring was required. Taken together, the results of the two experiments are clearly inconsistent with the hypothesis of a significant role for the direction of endogenous visual attention in the determination of ventriloquism.

Endogenous attention would only have provided an explanation for ventriloquism for those cases in which instructions require monitoring the visual attractor itself (e.g., Bertelson & Aschersleben, 1998) or in which the subject does so spontaneously. At first sight, a better case could be made for exogenous attention. There is good evidence (reviewed by Egeth & Yantis, 1997) that attention can be captured by visual stimuli with an abrupt onset, and the visual attractors used in ventriloquism experiments usually possess that feature. It is thus reasonable to inquire whether capture by exogenous attention is a condition of ventriloquism.

To investigate this question, a visual display was used that was thought capable of producing diverging effects on ventriloquism and on exogenous visual attention. This choice was guided by earlier data showing that exogenous attention can be captured by a visual item differing substantially by one or several attributes (such as color, form, orientation, or shape) from a set of identical items among which it is displayed (e.g., Treisman & Gelade, 1980). Such a unique item has been called a *singleton*. If ventriloquism is mediated by exogenous attention, one would predict that presenting a sound in synchrony with a display that contains a singleton will shift the apparent location of the sound toward that singleton. Consequently, finding a singleton that would *not* attract sound location, despite effectively capturing visual attention, would demonstrate a dissociation between exogenous attention and ventriloquism.

## EXPERIMENT 1

The critical idea was to use a singleton whose uniqueness consisted of being *smaller* than the other items in the display. The visual display consisted of a horizontal row of four bright squares, two on either side of a median fixation point, with a square smaller than the three other ones, situated at either the extreme left or the extreme right position (see Figure 1), as the singleton. In our endogenous attention study (Bertelson et al., 2000), we found, as has already been mentioned, that a display with a big square on one side of fixation and a small square on the other side shifted the reported origin of a centrally delivered sound toward the bigger square. This result suggested that our display might similarly attract a sound in the direction of the two big squares and away from the singleton. The attentional hypothesis, however, predicts that the visual display would bias the apparent sound location toward the singleton.

To examine that question, we used a new method introduced recently by Bertelson and Aschersleben (1998), in which the perceived location of target sounds is measured by using *psychophysical staircases*. In these authors' experiments, sounds whose location was controlled stereophonically were presented alone or in synchrony with flashes in a centrally located lamp. The task was to press one of two keys according to whether the sounds appeared to come from the left or the right of the center. Following the principle of psychophysical staircases, the successive locations of the sounds depended on the subject's responses: after a "left" response, the location of the following sound on the same staircase was moved one step to the right, and vice versa. For each condition, two staircases, one starting at the extreme left location, the other at the extreme right, were used in random alternation. Initially, the correct responses were given repeatedly, so that the staircases progressed monotonically toward the center. Then, at some point, *response reversals*, defined as responses different from the preceding one on the same staircase, began to occur. The main finding was that reversals occurred earlier at larger interstaircase distances with the central flashes than at those without them. This difference was attributed to the attraction of the sounds by the flashes, which brought their perceived location closer to the center than the objective location. This use of psychophysical staircases is different from habitual ones, in which the interest lies in the asymptote reached by each staircase, whereas here it lies in the locations at which the occurrence of response reversals shows that the subject does not know anymore where, relative to the center, the sounds came from.

In the present experiment, the staircase method was applied with the previously described visual display, with a small singleton either to the left or to the right of the display. Two staircases (one starting far left, the other far right) were run with each of the two displays (singleton left or right). A visual bias dependent on the position of the singleton should manifest itself at the level of the locations at which reversals begin to occur on the staircases for the two visual displays. If the apparent location of the sound is attracted toward the singleton, reversals should, on both staircases, first occur at locations more to the side *opposite* the singleton.

## Method

**Subjects.** Twelve students from Tilburg University received course credit for their participation. None of them reported any hearing or seeing problems.

**Experimental situation.** The subjects sat in front of a 15-in. computer screen (Olivetti DSM 60-510) at a distance of 50 cm, with their heads in a chinrest. Two loudspeakers (Philips box 410 car speakers) were positioned 50 cm on either side of the middle and at the same vertical level as the visual display. On each trial, three 100-msec tones of 2000-Hz sinewave sounds with no fade-in or fade-out were generated at 44.1 kHz (as in Bertelson et al., 2000). Their intensity was 53 dBa when measured at ear position (General Radio Sound Level Meter, Model 1563 in fast mode). The three tones were separated by 800-msec silent intervals.

The tones were accompanied by the presentation of a display ( $640 \times 480$  pixels) on the computer screen. The duration of the display was, like that of the tones, 100 msec (or seven cycles on a screen with a refresh rate of 70 Hz), and it flashed on and off in synchrony with the tones. A display consisted of four white squares against a dark background. There were three big squares of  $2.4^\circ$  ( $2.1 \times 2.1$  cm) and a small square of  $0.9^\circ$  ( $0.8 \times 0.8$  cm). The centers of the squares were horizontally aligned and displayed at eye level. The centers were each  $7.1^\circ$  apart (6.2 cm), so that the two outer squares were  $10.7^\circ$  from the middle and the two inner squares  $3.6^\circ$ . Two different displays were used, one with the small square on the far left, the other with the small square on the far right. A central fixation mark (+) appeared 800 msec before trial onset and remained visible for the duration of the trial.

**Procedure.** Testing was organized in three explorations. In each of them, successive trials were selected from four different randomly intermingled staircases. For each of the two displays (singleton on the left or the right), one staircase started with the sound's coming from an extreme left position, and the other started from an extreme right position. The apparent azimuthal origin of the sound was manipulated by varying the emission time of the sounds between the two loudspeakers (i.e., the interaural time difference). One phase-difference unit was equal to one sampling point, which, at a sampling rate of 44.1 kHz, is approximately  $22.6 \mu\text{sec}$ . This corresponds to about  $2.2^\circ$  of arc, if one assumes that the distance between the ears is 20 cm and that the sound travels directly to the left and right ears with a speed of 340 m/sec. When the left key was pressed, the following sound on the same staircase was moved one step to the right, and vice versa after a right keypress. The left staircase started with the left loudspeaker's leading by 15 phase-difference units, the right staircase with the opposite arrangement.

For each trial, the staircase controlling the locations of the auditory target and the visual display was chosen at random. An exploration was stopped as soon as nine reversals had been recorded on each of the four staircases. Each trial started 2 sec after the response to the preceding one. An exploration generally lasted about 12–15 min, after which there was a small pause. The whole session lasted about 45 min.

The subjects were instructed to indicate, by pressing one of two keys on a special keyboard, whether the sound had come from the left or from the right of the fixation point. They were asked to keep their gaze on the fixation point for the duration of each trial.

## Results

Mean locations at which response reversals occurred are shown in Figure 2, separately for each of the four staircases. As was noted by Bertelson and Aschersleben (1998), the sawtooth appearance of the curves is an unavoidable consequence of the method, because successive reversals are necessarily separated by at least one phase difference step. The important aspect of the results is that staircases were shifted in the *direction of the singleton*. This indicates that the apparent origin of the sound was shifted not toward the singleton, but instead in the opposite direction, because, as was explained in the introduction, when the apparent location of the sounds is attracted by one particular visual event, response reversals occur *farther away* from that event.

The statistical analysis was based on the locations at which response reversals occurred. The mean overall distance between these locations for the two staircases with a singleton on the left versus the right was  $20.7 \mu\text{sec}$  (equivalent to approximately  $2.0^\circ$ ). By paired *t* test, this

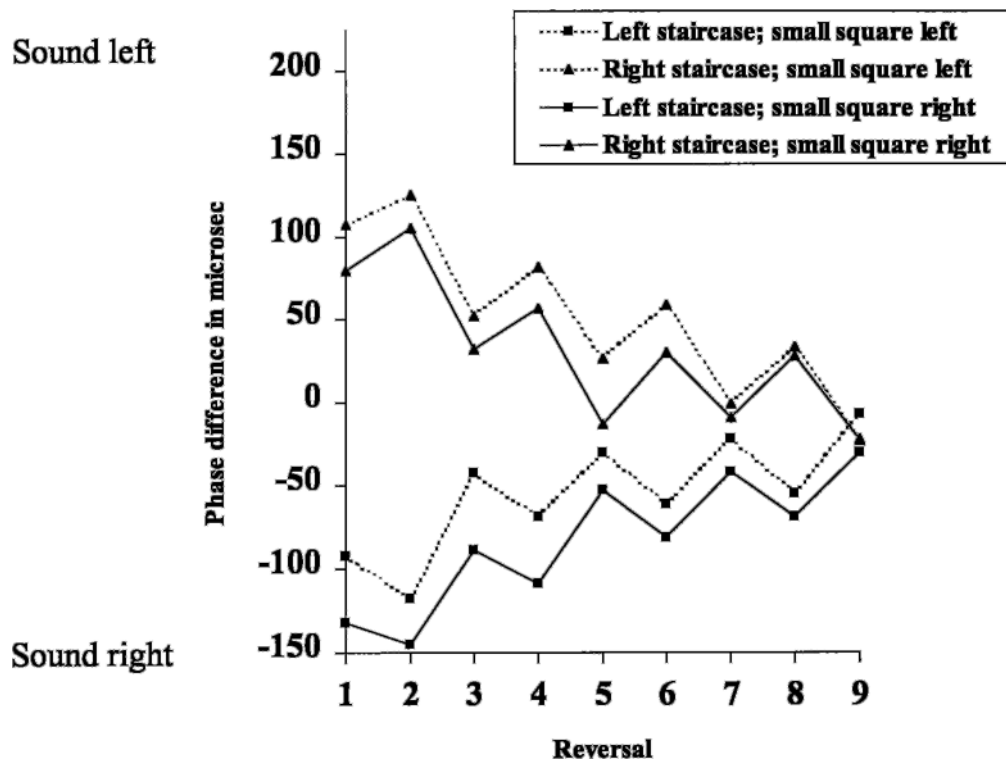


Figure 2. Mean location, in phase-difference units, of the first nine response reversals on each staircase in Experiment 1.

difference was significant [ $t(11) = 3.38, p < .006$ , two-tailed]. Inspection of the individual data showed that the sound was shifted toward the two big squares for each of the 12 subjects ( $p < .001$  by sign-test).

### Discussion

The results are very straightforward. The apparent origin of the sound was not shifted toward the singleton, but actually in the opposite direction—that is, toward the two big squares on the other side of the display. Presumably, these two big squares were attracting the apparent origin of the sounds more strongly than were the small and big squares at the other side. Thus, the attractor size effect, previously obtained by Bertelson et al. (2000), occurred with the present visual display also.

The fact that the singleton did not produce ventriloquism seems inconsistent with the attention capture explanation of that effect. However, before that conclusion is fully accepted, it is necessary to verify that our visual display produces the expected effect on exogenous attention. It is known, actually, that the mere presence of a singleton may not be enough to capture attention (Jonides & Yantis, 1988). Moreover, we are not aware of any study showing that a singleton captures exogenous attention when its uniqueness consists of being smaller (or, e.g., less intense) than the other items in the display. The next experiment was run to determine whether our singleton effectively captured exogenous attention.

### EXPERIMENT 2

The principle of this experiment was to measure the attention attraction capacity of the small singleton through its effect on the discrimination of targets presented elsewhere in the display. If a singleton effectively captures attention, it will hamper processing of other items in the display. This approach has already been used by Pashler (1988) and by Theeuwes (1991), and it is now called the *additional singleton paradigm* (for a review, see Simons, 2000). The subjects were presented with either the previously used display, with three big squares and a small one to the left or the right (singleton condition), or one with four identical big squares (control condition). In both conditions, a target letter, calling for a two-choice reaction, was superimposed on either the extreme left or the extreme right square. If the small square captures attention, one would expect discrimination of the target letter to be slower when presented on the big square opposite the singleton in the singleton condition than when presented on a big square in the corresponding location in the control condition. For the trials with the target on the singleton, there was no testable prediction, because a comparison with the control condition would have involved targets located on squares of unequal sizes. Proximity with the contours might, for example, slow down discrimination of targets presented on the smaller squares. For these reasons, data from trials with the tar-

get on the singleton were not used to assess the attention-capturing capacity of the singleton. It was nevertheless necessary to have 50% of such filler trials in the singleton condition, because in their absence, the position of the singleton would have provided a valid cue regarding target location that was not available in the control condition.

### Method

**Subjects.** Fourteen new subjects received course credit for their participation. All had normal or corrected-to-normal vision.

**Materials.** The display for the singleton condition was the one already used in Experiment 1, with three big squares and one small singleton. For the control condition, the singleton was replaced by a big square, so that the display now contained four equally sized big squares. In both conditions, a target letter, either X or O (Arial bold, size 10, color black) was presented on every trial in the center of either the extreme left or the extreme right square.

**Procedure.** The subjects sat, as before, in front of a computer screen at a distance of 50 cm and with their heads in a chinrest. On each trial, a central fixation point appeared 800 msec before a 100-msec presentation of the display and stayed on until the latter's offset. The subjects had to press the left key for letter X and the right one for letter O. Twelve different stimuli were created by the orthogonal combination of trial type (singleton left, singleton right, or no singleton), target letter (X or O), and target position (left or right). Each of the 12 different stimuli was presented 12 times in a single block of 144 trials. All the stimuli were presented in random order. Twenty-four practice trials were run before testing proper.

The subjects were instructed to keep their eyes fixated on the central dot. They were informed that the target letter could appear equally often in the extreme left or the extreme right square of the display and that, in case there was a small square, the target letter appeared equally often in that square or in the big square at the other end. Corrective feedback was provided after each trial. The inter-trial interval was 3.5 sec.

### Results and Discussion

Reaction times (RTs) shorter or longer than 2.5 standard deviations from the individual mean were discarded. They represented 2.2% of the trials. For reasons already explained, in the singleton condition, only the trials with the target letter on the big square opposite the singleton were analyzed. More errors were made on these trials (26.7%) than in the control condition (22.0%), a difference that was significant [ $t(13) = 3.31, p < .003$ , one-tailed; the use of one-tailed testing is justified because there was a clear prediction]. RTs were also longer in the singleton condition (752 msec) than in the control condition [737 msec;  $t(13) = 1.85, p < .04$ ]. Thus, both speed and accuracy of target discrimination were worse in the singleton condition than in the control condition. Apparently, then, the small singleton effectively attracted exogenous attention away from the target letter on the opposite side of the display.

Combined with the finding in Experiment 1 that apparent sound location was biased not toward the singleton, but in the opposite direction, this result would seem to show a dissociation between attentional capture and ventriloquism. However, before that conclusion can be fully accepted, the possibility must be considered that the capture of attention by the singleton demonstrated in Experiment 2 did not occur in Experiment 1 because of

differences between the situations of the two experiments. Specifically, in Experiment 2, the task of having to identify a target letter may have created a voluntary set favorable to attentional capture, which was not called for in Experiment 1. It has indeed been shown that attentional capture by feature singletons depends on the expectations of the subject (e.g., Egeth & Yantis, 1997). A new experiment was therefore run, in which we tried to measure the two effects of our singleton under equalized expectation conditions.

## EXPERIMENT 3

In this experiment, the effects of the smaller singleton on attention capture and the apparent location of a simultaneously occurring sound were measured on separate trials of the same experiment, with subjects not knowing in advance whether they would have to localize the sound or discriminate target letters.

### Method

**Subjects.** Twelve new subjects received course credit for their participation. All had normal or corrected-to-normal vision.

**Materials.** Two types of trials were presented. On *localization trials*, the task was, just as in Experiment 1, to localize the sound; on *discrimination trials*, it was, just as in Experiment 2, to identify the target letter as X or O. On both types of trials, the stimuli were now presented once for 160 msec. The visual display for sound localization consisted, as in Experiment 1, of three big squares and one small square on the extreme left or right synchronized with a tone controlled by one of the four staircases. The visual display for discrimination trials consisted, as in Experiment 2, of either four big squares (control condition) or three big squares and a small square on the left or right (singleton condition), with a target letter on either the extreme left or the extreme right square. Unlike Experiment 2, discrimination trials were now accompanied by a tone. The tone was delivered in the same location as on the immediately preceding localization trial.

**Procedure.** The whole experiment consisted of four blocks of 180 trials each. Each block comprised 120 sound localization trials (30 trials for each of the four staircases) mixed randomly with 60 X/O discrimination trials (five repetitions of the 12 different stimuli created by the orthogonal combination of trial type [singleton left, singleton right, or no singleton], target letter [X or O], and target position [left or right]).

The subjects were instructed to keep their eyes fixated on the central dot that preceded display onset. They were to localize the sound as coming from the left or the right by pressing a left or right key. Whenever an X or an O was displayed in one of the outer squares, they had to press, as quickly and accurately as possible, the left key for letter X and the right key for letter O. To avoid subject's confusing sound localization trials with X/O discrimination trials (because they had missed the target letter), discrimination trials were immediately followed by the letter X or O, with arrows pointing toward the appropriate keys ( $\leftarrow$ X or O $\rightarrow$ ). Testing lasted about 1 h, and the first block was treated as practice.

### Results and Discussion

**Sound localization.** Data were analyzed as in Experiment 1. The mean locations at which the first nine response reversals occurred are shown in Figure 3, separately for each of the four staircases. As is apparent, staircases were, as before, shifted in the direction of the

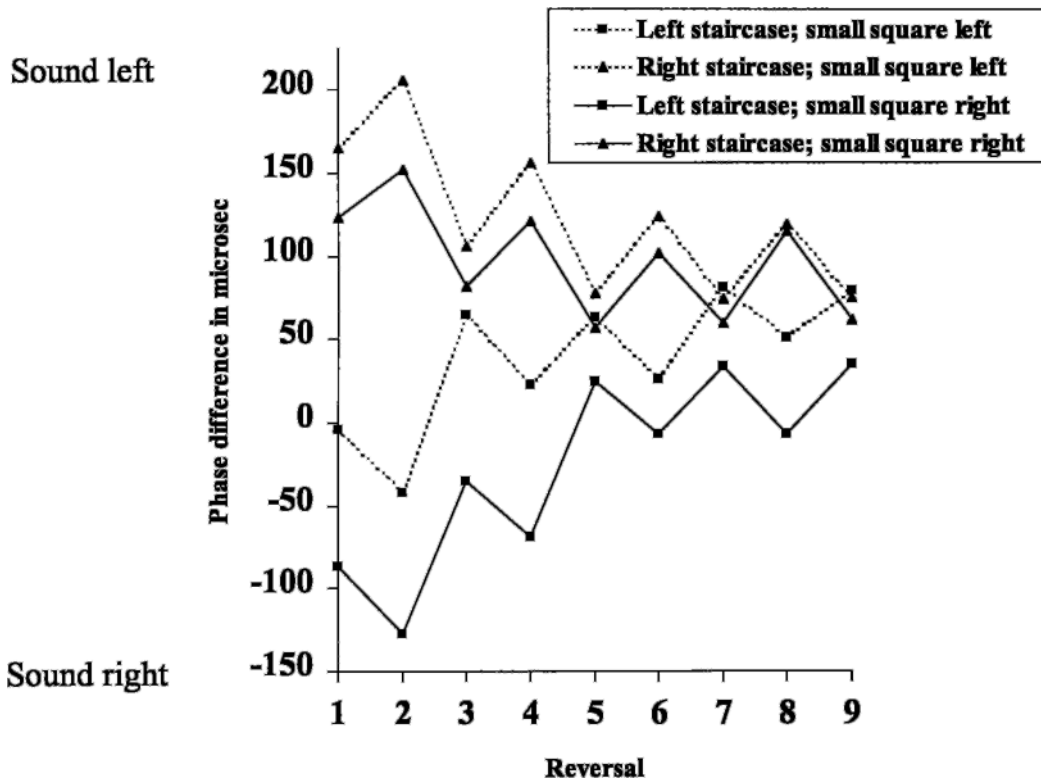


Figure 3. Mean location, in phase-difference units, of the first nine response reversals on each staircase in Experiment 3.

singleton, indicating that the apparent origin of the sound was shifted not toward the singleton, but toward the two big squares in the opposite direction. The mean overall distance between the locations for the two staircases with a singleton on the left versus the right was  $89.9 \mu\text{sec}$  (equivalent to approximately  $9.0^\circ$ ). By paired  $t$  test, this difference was significant [ $t(11) = 3.32, p < .007$ , two-tailed]. Inspection of the individual data showed that the sound was shifted toward the two big squares for 11 out of 12 subjects ( $p < .005$  by sign-test).

**X/O discrimination.** RTs and errors were analyzed as in Experiment 2. As before, more errors were made in the singleton condition (24.1%) than in the control condition (15.8%), a difference that was significant [ $t(11) = 4.35, p < .001$ , one-tailed]. Eleven out of 12 subjects made more errors in the singleton condition than in the control condition ( $p < .005$ ). RTs were also longer in the singleton condition (975 msec) than in the control condition [939 msec;  $t(11) = 2.53, p < .02$ ], with 10 out of 12 subjects being slower in the singleton condition ( $p < .025$ ). Thus, both speed and accuracy of target discrimination were worse in the singleton condition than in the control condition, indicating that the singleton effectively attracted exogenous attention away from the target letter at the opposite side of the display.

Experiment 3 thus replicates and extends the results of our previous experiments: Exogenous attention was shifted toward the singleton, whereas the apparent origin of the sound was shifted toward the two big squares in the opposite direction. Both effects were now measured under the same conditions within the same experiment, thus excluding potential strategic confounds.

## GENERAL DISCUSSION

The general goal of the study was to test whether the spatial biasing of an apparent sound origin by a spatially discordant visual attractor is conditional upon capture of exogenous attention by the attractor. The results allow an unequivocally negative answer. In Experiment 1, when sound bursts were presented in synchrony with the visual display with the smaller singleton on one side, the apparent location of the sounds was shifted not toward the singleton, but instead away from it and toward the big squares at the other end of the display. Before conclusions could be drawn from that result concerning the role of attention in ventriloquism, it was necessary to verify that the singleton in our display effectively captured exogenous visual attention. In Experiment 2, occurrence of that singleton deteriorated the discrimination of targets pre-

sented at the other end of the display, showing that it effectively attracted exogenous attention. In Experiment 3, we verified that strategic differences between the two experiments could not account for the contrasting results. Taken together, the experiments thus demonstrate a dissociation between exogenous attention, which is attracted by the singleton, and apparent sound location, which is not. That pattern of results is inconsistent with the eventual mediating role of the direction of exogenous attention in the generation of ventriloquist effects.

The findings from the present study concerning the role of *exogenous* attention, together with those of the earlier one showing the independence of ventriloquism from the direction of *endogenous* attention (Bertelson et al., 2000), support the general conclusion that ventriloquism is not affected by the direction of attention. As has already been suggested in the previous paper, the interactions at the root of ventriloquism and of other cross-modal effects probably take place at a stage of processing *before* attentional selection. This stage is presumably one concerned with the initial analysis of the spatial scene (Bertelson, 1994). This presumption receives additional support from the already mentioned finding by Driver (1996) that ventriloquism can improve selective attention to a particular message in a multichannel listening situation, implying that the selection of an attentional focus operates on a representation of the external scene that has already been spatially reorganized by cross-modal interactions. A recent study of ours (Vroomen & de Gelder, 2000), in which we showed that the perceptual organization of a sound affects visual scene analysis, also supports that way of seeing things.

Apparently, there is now a clear answer to the question asked at the start of the two studies: The direction of visual spatial attention, either endogenous or exogenous, does not provide an alternative interpretation to accounts of ventriloquism in terms of automatic interactions between sensory codes for location, eventually combined with more cognitive influences.

It might be objected that our study dealt with only one determinant of the direction of exogenous visual attention, the presence of a singleton, and that different results might be obtained by manipulating another determinant, the abruptness of the visual attractor's onset. We noted already that the visual attractors used in most studies of ventriloquism have abrupt onsets; hence, the question of a possible mediation of ventriloquism might be asked again about this factor. There are however two reasons, already mentioned in an earlier paper (Bertelson et al., 2000), to doubt that capture of exogenous attention by abrupt onsets would, any more than capture by a singleton, affect ventriloquism. First, abrupt onset visual attractors that are not synchronized with the target sounds produce no ventriloquism, or produce less than do synchronized ones (Bertelson & Aschersleben, 1998; Bertelson, Vroomen, & de Gelder, 1997; Choe et al., 1975; Radeau & Bertelson, 1977, 1987), yet must attract exogenous attention to the same extent. Second, in Bertelson et al.'s (2000) experiments, full ventriloquist effects were obtained with endogenous attention directed *away* from

the visual attractor, a condition under which Yantis and Jonides (1990) have shown that exogenous capture no longer occurs.

The fact that the direction of spatial attention does not influence the size of ventriloquist effects may seem difficult to reconcile with existing evidence for a modulation of cross-modal effects by another attentional factor, selective *attention to modality*. For instance, in experiments on AV cross-modal bias, subjects do not report the same location when instructed to localize either the visual or the auditory stimulation (Bertelson & Radeau, 1981; Radeau, 1992). Thus, modality selection, even if not perfect, is at least partially effective. It has been shown also, in studies of the aftereffects of exposure to AV conflict, that the respective sizes of auditory versus visual aftereffects can be modified by having the subject localize inputs in one or the other modality during the exposure period (Canon, 1970; Radeau, 1974). It has been suggested elsewhere (Bertelson, 1999) that the implications of these facts for theories of cross-modal interaction depend, just as is the case for the effects of cognitive determinants, on the stage of processing at which they originate. If it were to turn out that this stage was a relatively late one, like the attribution of weights to the outputs from the respective modalities in the readout process leading to the localization response, there would be no contradiction with the conclusions from work with directional attention.

In laboratory situations such as the present one, ventriloquism leads to mislocalization of auditory sources. However, in the real world, auditory sources typically will correspond in location to matching visual events, rather than being discrepant. Hence, the cross-modal interactions that we have been studying should normally be adaptive, tending to favor veridical rather than illusory spatial perception. With regard to endogenous attention, we have argued that from a functional perspective, it makes sense that ventriloquism is not affected by where the subject is focusing attention. If auditory localization were affected by wherever a subject chose to attend to visually, the apparent location of a fixed auditory source would change every time the person decided to shift his visual attention. A similar case can now be made for exogenous attention. If auditory localization changed whenever a visual stimulus captured exogenous attention, it would presumably be impossible to keep an internal representation of space consistent with external reality. It thus seems better that cross-modal spatial interactions should be driven by stimulus factors in a bottom-up fashion, rather than being susceptible to wherever attention is focused or captured. The finding that ventriloquism can be dissociated from attention thus makes good functional sense.

Apart from its role in bringing about the dissociation of ventriloquism from exogenous attention, our finding that attention can be captured by a singleton that is smaller than other items in the visual field is an important addition to current knowledge about the mechanisms of attention. As far as we know, this has never been shown. One of its implications is that the frequent characteriza-



tion of attentional capture as “attraction toward salient stimuli” can be misleading, if “salient” is taken as meaning “bigger” or “brighter.”

## REFERENCES

- BERMANT, R. I., & WELCH, R. B. (1976). The effect of degree of visual-auditory stimulus separation and eye position upon the spatial interaction of vision and audition. *Perceptual & Motor Skill*, **43**, 487-493.
- BERTELSON, P. (1994). The cognitive architecture behind auditory-visual interaction in scene analysis and speech identification. *Current Psychology of Cognition*, **13**, 69-75.
- BERTELSON, P. (1999). Ventriloquism: A case of cross-modal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347-362). Amsterdam: Elsevier.
- BERTELSON, P., & ASCHERSLEBEN, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, **5**, 482-489.
- BERTELSON, P., & RADEAU, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, **29**, 578-584.
- BERTELSON, P., VROOMEN, J., & DE GELDER, B. (1997). Auditory-visual interaction in voice localization and bimodal speech recognition: The effect of desynchronization. In C. Benoit & R. Campbell (Eds.), *Proceedings of the Workshop on Audio-Visual Speech Processing: Cognitive and computational approaches* (pp. 97-100). Rhodes, Greece.
- 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.
- 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.
- 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.
- 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. J. (1994). Cross-modal synergies in attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 311-331). Cambridge, MA: MIT Press, Bradford Books.
- EGETH, H. E., & YANTIS, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, **48**, 269-297.
- HELD, R. (1965). Plasticity in sensorimotor systems. *Scientific American*, **213** (5), 84-94.
- JACK, C. E., & THURLOW, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual & Motor Skills*, **38**, 967-979.
- JACKSON, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, **5**, 52-65.
- JONIDES, J., & YANTIS, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, **43**, 346-354.
- PASHLER, H. (1988). Cross-dimensional interaction and texture segregation. *Perception & Psychophysics*, **43**, 307-318.
- PICK, H. L., JR., WARREN, D. H., & HAY, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, **6**, 203-205.
- POSNER, M. I., & COHEN, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of language processes* (pp. 531-555). Hillsdale, NJ: Erlbaum.
- RADEAU, M. (1974). Adaptation au déplacement prismatique sur la base d’une discordance entre la vision et l’audition [Adaptation to prismatic displacement on the basis of discordance between vision and audition]. *L’Année Psychologique*, **74**, 23-34.
- RADEAU, M. (1985). Signal intensity, task context, and auditory-visual interaction. *Perception*, **14**, 571-577.
- RADEAU, M. (1992). Cognitive impenetrability in auditory-visual interaction. In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 41-55). Amsterdam: Elsevier.
- RADEAU, M., & BERTELSON, P. (1974). The after-effects 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. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, **22**, 137-146.
- RADEAU, M., & BERTELSON, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*, **23**, 341-343.
- RADEAU, M., & BERTELSON, P. (1987). Auditory-visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychological Research*, **49**, 17-22.
- SIMONS, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, **4**, 147-155.
- SPENCE, C., & DRIVER, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1005-1030.
- SPENCE, C., & DRIVER, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.
- THEEUWES, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, **50**, 184-193.
- TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **5**, 109-137.
- VROOMEN, J. (1999). Ventriloquism and the nature of the unity assumption. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 389-393). Amsterdam: Elsevier.
- VROOMEN, J., & DE GELDER, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1583-1590.
- WARD, L. (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., WELCH, R. B., & MCCARTHY, T. J. (1981). The role of visual-auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, **30**, 557-564.
- WELCH, R. B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception & Psychophysics*, **12**, 453-456.
- WELCH, R. B. (1978). *Perceptual modification: Adaptation to altered sensory environments*. New York: Academic Press.
- 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: Elsevier.
- YANTIS, S., & JONIDES, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 121-134.