Research Report

VISUAL MOTION INFLUENCES THE CONTINGENT AUDITORY MOTION AFTEREFFECT

Jean Vroomen and Beatrice de Gelder

Tilburg University, Tilburg, the Netherlands

Abstract—*In this study, we show that the contingent auditory motion aftereffect is strongly influenced by visual motion information. During an induction phase, participants listened to rightward-moving sounds with falling pitch alternated with leftward-moving sounds with rising pitch (or vice versa). Auditory aftereffects (i.e., a shift in the psychometric function for unimodal auditory motion perception) were bigger when a visual stimulus moved in the same direction as the sound than when no visual stimulus was presented. When the visual stimulus moved in the opposite direction, aftereffects were reversed and thus became contingent upon visual motion. When visual motion was combined with a stationary sound, no aftereffect was observed. These findings indicate that there are strong perceptual links between the visual and auditory motion-processing systems.*

In psychophysical studies on the interaction of information in the visual and auditory modalities, localization of auditory stimuli has often been investigated. In the classical ventriloquist illusion, in which a static light and static sound are presented simultaneously but at different positions, participants perceive the position of the sound as shifted toward that of the light (for reviews, see Bertelson, 1999; Vroomen & de Gelder, in press). The ventriloquist effect is very robust and can be observed even when participants are explicitly trained to ignore the visual distractor (Vroomen, Bertelson, & de Gelder, 1998); when cognitive strategies of the participant can be excluded (Bertelson & Aschersleben, 1998); when the visual distractor does not command attention, either endogenously (Bertelson, Vroomen, & de Gelder, 2000) or exogenously (Vroomen, Bertelson, & de Gelder, 2001b); and when the visual stimulus goes undetected in one hemifield as a consequence of contralateral brain damage, as in the case of hemineglect (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000).

Investigations of cross-modal spatial interactions with dynamic events are, however, far less frequent, and at present generalizations from static to dynamic events may not be warranted, as it has been argued that position and motion of an object are processed independently (e.g., Whitney, 2002). Yet it has been observed that a light stimulus moving on a frontal screen creates the impression that a static sound is moving as well, a phenomenon Mateeff, Hohnsbein, and Noack (1985) called *dynamic visual capture*. They reported that the faster the light stimulus moved, the stronger the motion of the sound was sensed. Kitajima and Yamashita (1999) observed a similar effect in that the perceived direction of a moving sound was greatly influenced by the direction in which a light stimulus moved, even when the physical motion directions of the stimuli were perpendicular or opposite to each other. Soto-Faraco and Kingstone (in press) also reported that visual apparent motion can have a strong effect on auditory apparent motion.

These results suggest the existence of strong cross-modal links in the perception of dynamic events, in the sense that visual motion affects auditory motion perception. However, these dynamic-visual-capture effects might also reflect some form of confusion at the level of the instructions given to participants or a response bias elicited by the experimental situation, rather than a genuine perceptual phenomenon. Several authors have pointed out that intersensory conflict situations like the one used in cross-modal capture are open to alternative explanations, in particular if participants are aware of the conflict or if the situation is transparent (Bertelson, 1999; Vroomen, Bertelson, & de Gelder, 2001a; Welch, 1999). For example, participants may occasionally report, despite instructions not to do so, the direction of the to-be-ignored visual stimulus rather than the direction of the target sound. If so, then at least part of the visual-capture phenomenon could be attributed to confusion between target and distractor modality. Therefore, conscious strategies, confusion, or response-related decision processes need to be dissociated from true cross-modal perceptual effects if behavioral phenomena like dynamic visual capture are to be attributed to the latter.

Our goal was to study dynamic audiovisual events by creating a situation that was not transparent to the participants. We took advantage of a recently discovered phenomenon described as the contingent auditory motion aftereffect (CAMA; Dong, Swindale, & Cynader, 1999). It is an auditory analogue of the visual contingent aftereffect known as the McCollough effect (McCollough, 1965). Dong et al. observed that after about 10 min of induction, during which subjects listened to a rightward-moving sound with a falling pitch alternated with a leftward-moving sound with a rising pitch, a stationary sound with a falling pitch was perceived as moving leftward, whereas a stationary sound with a rising pitch was perceived as moving rightward.

We first examined whether this auditory phenomenon could be modified by the introduction of a visual stimulus that moved during the induction phase either in the same direction as the sound or in the opposite direction. If there is a link between the auditory and visual motion-processing systems, one might expect the CAMA to be reduced or even reversed when visual and auditory motion are in opposite directions. Similarly, when visual and auditory motions are in the same direction, one might expect an enhancement of the CAMA compared with the effect observed after an auditory-only induction phase.

Second, we examined whether a stationary sound combined with a mobile visual stimulus could induce a CAMA as well. If a mobile visual stimulus induces motion of a stationary sound (i.e., visual capture), one might expect aftereffects in this condition to be contingent upon visual motion. Moreover, the stationary-sound condition could also serve as a control to check whether fixating a mobile visual stimulus with the eyes has any additional effect on the CAMA. Alternatively, though, if a cross-modal effect reflects a direct interaction between the auditory and visual motion-processing systems, one might

Address correspondence to Jean Vroomen, Tilburg University, Department of Psychology, Warandelaan 2, Tilburg, The Netherlands; e-mail: j.vroomen@kub.nl.

not obtain a CAMA when a stationary sound is combined with a moving visual stimulus, as a stationary sound may not activate the auditory motion-processing system.

METHOD

Participants

A total of 40 subjects participated. They were randomly assigned to an audiovisual congruent condition, an auditory-only condition, an audiovisual incongruent condition, or a stationary-sound condition (10 participants each). Participants were all first-year psychology students from Tilburg University, Tilburg, the Netherlands. They reported normal or corrected-to-normal vision and normal hearing. All were unaware of the purpose of the experiment.

Apparatus

Testing was administered in a darkened, sound-insulated cubicle. Participants sat in front of a 15-in. computer screen positioned at eye level at a distance of 90 cm from the eyes. Two small loudspeakers were placed in front of the screen, one 84 cm from each side. The separation between the centers of the cones of the loudspeakers was 32 cm (or 22°).

Stimuli and Design

Auditory stimuli

The auditory stimuli were modeled after those used by Dong et al. (1999). All sounds were created with Cooledit 2000 (Syntrillium Software Corp., Phoenix, AZ) and stored as digital files on the hard disk of a personal computer. The sound was a 1-octave band-pass noise of 1 s duration. The center frequency of the band-pass noise was either rising or falling. For the sound with rising pitch, the center frequency started at 1177 Hz and then changed to 1912 Hz at 0.7 octaves/s. For the sound with falling pitch, the center frequency started at 1912 Hz and then changed to 1177 Hz at −0.7 octaves/s. Both sounds had a linear fade-in and fade-out of 25 ms each. They were delivered stereophonically through two loudspeakers connected to a SoundBlaster card (Soundblaster 16, Creative Labs, Singapore). The loudness of the sounds was 75 dBA when measured at ear level.

For the induction phase of the audiovisual congruent, auditory-only, and audiovisual incongruent conditions, auditory spatial motion was simulated by varying the intensity difference between the two loudspeakers. For leftward-moving sounds, the intensity of the sound on the left speaker started at 100% of the original intensity, and then decreased linearly to 0% in 1 s, while at the same time the intensity of the right speaker started at 0% and then increased linearly to 100%. The opposite arrangement was used for rightward-moving sounds. For the stationary-sound condition, a stationary sound was presented, with the intensities of the left and right speakers at 50% of the original intensity throughout.

For the test phase, sounds with ambiguous direction of motion were created. Two seven-step motion continua were created, one for sounds with rising pitch and one for sounds with falling pitch. The stimuli were the same as the induction sounds, except for the intensity of the left and right speakers. For the first stimulus of each continuum, the intensity varied from 20% to 80% of the original intensity. For each next stimulus on the continuum, the intensities at the onset and offset were changed by 10%. Thus, for the second stimulus with leftward motion, the intensity of the left speaker started at 70% and then decreased linearly to 30%, whereas the intensity of the right speaker started at 30% and then increased to 70%. The fourth stimulus in each continuum was stationary, as the intensities of the left and right speakers remained at 50% throughout.

Visual stimulus

The visual stimulus consisted of a bright white square that measured 2.1 × 2.1 cm and moved against the dark background of the computer screen. During 1 s, the square moved 22 cm (14°/s) in the horizontal plane from the far-left to the far-right position of the screen (or vice versa), such that its start and end positions were as close as possible to the cones of the loudspeakers.

Design and Procedure

The experiment consisted of three parts: a pretest, an induction phase, and a posttest. The *pretest* served as the baseline for computing aftereffects. A two-alternative forced-choice paradigm was used. Each of the 7 motion stimuli of the two continua was presented 12 times in quasi-random order, for a total of 168 experimental trials. Within each block of 28 successive trials, each of the 14 different stimuli appeared twice. Participants pressed a left key to indicate leftward motion, and a right key to indicate rightward motion. The next trial started 2 s after a key press. Six trials were given as practice before testing started.

The *induction phase* immediately followed the pretest. In the audiovisual congruent, auditory-only, and audiovisual incongruent conditions, leftward-moving sounds with rising pitch and rightward-moving sounds with falling pitch (or vice versa) were alternated systematically with a silent interval of 1,400 ms between two successive sounds. In the *audiovisual congruent* condition, participants heard the adapter sounds while watching the square moving in the same direction as the sounds. The square moved when a sound was presented, and remained stationary during the silent interval between two successive sounds. The *audiovisual incongruent* condition was the same, except that the square moved in the opposite direction. In the *auditory-only* condition, moving sounds were presented unimodally with no visual stimulus on the screen. In the *stationary-sound* condition, the stationary sound with rising pitch was alternated with the stationary sound with falling pitch while the visual stimulus moved from left to right and right to left.

In each of the four conditions, half of the participants were exposed to rightward-moving sounds with rising pitch alternated with leftward-moving sounds with falling pitch (referred to as rise-right/fall-left); for the other half of the participants, the induction schema was reversed (rise-left/fall-right). Induction lasted for 250 presentations of a single pair of leftward- and rightward-moving sounds (or 20 min). Participants in the congruent, incongruent, and stationary-sound conditions were instructed to fixate their gaze on the mobile visual stimulus that was presented during the induction phase. Thus, for 20 min they followed the left-to-right and right-to-left movement of the visual stimulus. The experimenter checked the eye movements of the participants during the entire induction phase.

The *posttest* started immediately after the induction period and was the same as the pretest, except that the test stimuli were presented in a different random order.

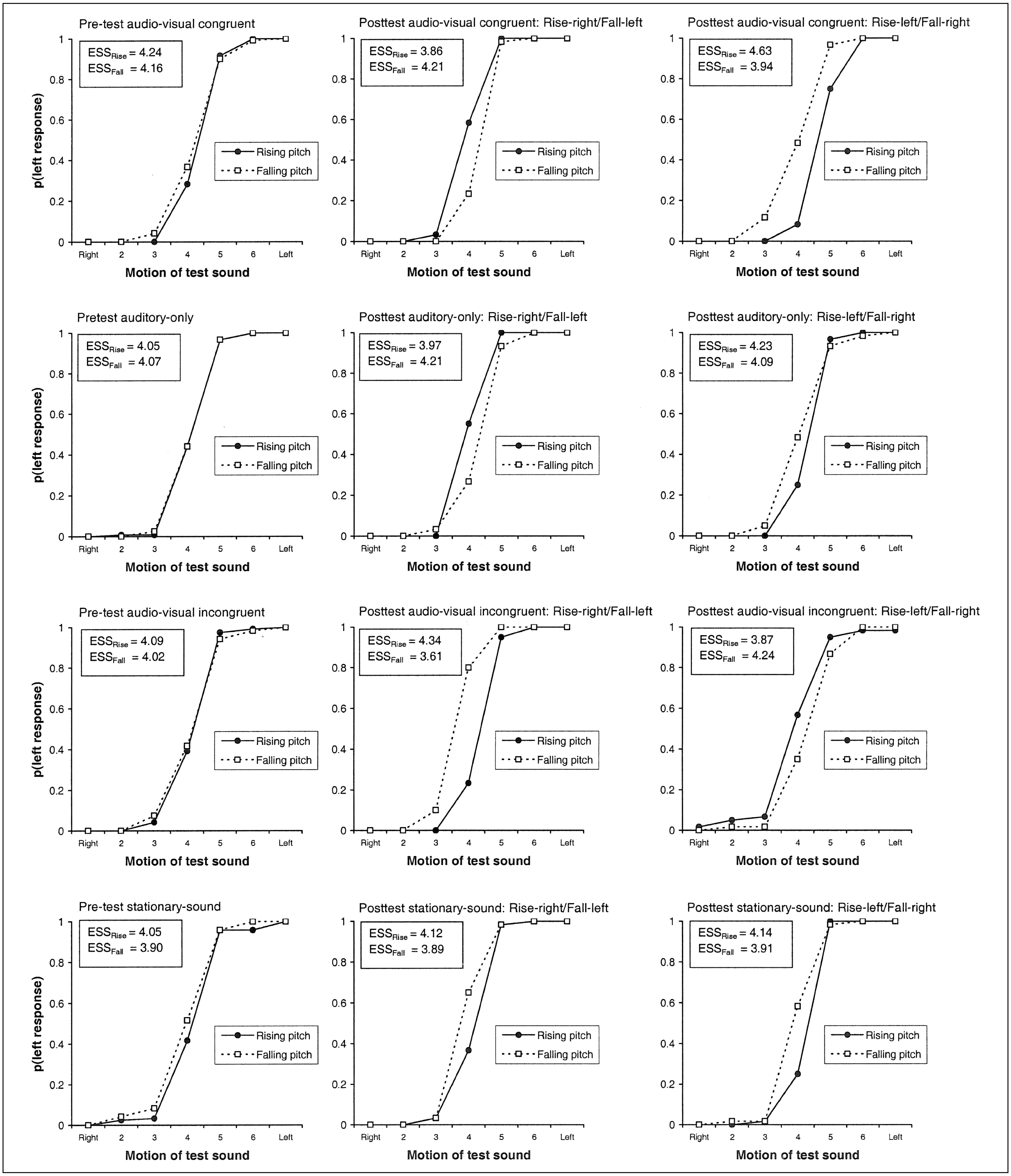

Fig. 1. Proportion of "left" responses as a function of auditory motion stimulus for sounds with rising and falling pitch, during pretest (left column) and posttest (center and right columns). In order from top to bottom, results are shown separately for the audiovisual congruent condition, auditory-only condition, audiovisual incongruent condition, and stationary-sound condition. The inset in each graph shows the estimated stationary stimulus (ESS) for sounds with rising pitch and sounds with falling pitch.

RESULTS

A psychometric function was obtained by plotting the percentage of "left" responses for each condition and each stimulus of the two motion continua (see Fig. 1). As expected, when participants were induced to the rise-right/fall-left schema in the audiovisual congruent condition (middle graph in the top row of the figure), a stationary sound (stimulus 4) with rising pitch was perceived as moving more leftward than a stationary sound with falling pitch. Conversely, when participants were induced to the rise-left/fall-right schema (right graph in the top row of the figure), a stationary sound with rising pitch was perceived as moving more rightward than a stationary sound with falling pitch. Smaller shifts, but in the same direction, were observed in the auditory-only condition. In the audiovisual incongruent condition, the shift was reversed, and there was no effect in the stationary-sound condition.

To determine the size of the aftereffects, we fitted a cumulative normal distribution for each participant so that we obtained an estimate of the stimulus that was perceived to be stationary. The average values of this *estimated stationary stimulus* (ESS) for sounds with rising and falling pitch are presented in the box inserts of Figure 1. Aftereffects were computed by determining, for each subject, the ESS for rising and falling pitches in the pretest and in the posttest. Then, the difference between rising and falling pitches in the pretest and between rising and falling pitches in the posttest was calculated. Finally, one of these two difference scores was subtracted from the other such that positive numbers referred to aftereffects in the predicted direction (see Table 1). For the stationary-sound condition, aftereffects were determined as if contingent upon visual motion.

In the overall analysis of variance on the aftereffects, there was a main effect of condition, $F(3, 40) = 11.87$, $p < .001$, but no effect of induction schema and no interaction between induction schema and condition (both Fs < 1). Aftereffects were therefore pooled across induction schemas. Aftereffects in the audiovisual congruent condition were bigger than those in the auditory-only condition, $t(18) = 2.30$, $p < .02$; audiovisual incongruent condition, $t(18) = 4.99$, $p < .001$; and stationary-sound condition, $t(18) = 3.43$, $p < .002$ (all tests one-tailed). Also, aftereffects in the auditory-only condition were bigger than those in the audiovisual incongruent condition, $t(18) = 3.81$, $p < .001$, and aftereffects in the audiovisual incongruent condition were smaller than those in the stationary-sound condition, $t(18) = 2.95$, $p < .004$.

Separate t tests showed that aftereffects in the audiovisual congruent and auditory-only conditions were significantly larger than chance level, $t(9) = 3.75$, $p < .003$, and $t(9) = 1.91$, $p < .05$, respectively. Aftereffects in the audiovisual incongruent condition were significantly smaller than chance level, $t(9) = -3.46$, $p < .004$, indicating that aftereffects were not contingent upon auditory, but upon visual motion. Aftereffects in the stationary-sound condition were not significantly different from chance level, $t(9) = 0.90$, n.s.

DISCUSSION

The present study shows that the CAMA as originally reported by Dong et al. (1999) is strongly affected by visual motion information. Auditory aftereffects were bigger when a visual stimulus moved in the same direction as a sound during the induction phase than when no visual stimulus was presented during induction. When the visual and auditory stimuli moved in opposite directions, aftereffects were reversed and contingent upon visual motion. These data thus clearly indicate that visual motion affects auditory motion processing.

Other studies have also reported visual effects on auditory motion processing (Kitajama & Yamashita, 1999; Mateeff et al., 1985; Soto-Faraco & Kingstone, in press), but they failed to truly control for cognitive biases. The present approach was successful, though, because the situation was not transparent to the participants. That is, even though they may have noticed an intersensory conflict during the induction phase, it seems impossible that this knowledge could have guided strategies or hypotheses about how to respond in the unimodal auditory test phase.

In addition, we observed that a stationary sound combined with a mobile visual stimulus did not produce a CAMA. One might have expected that the visual stimulus would capture the apparent location of the stationary sound (Kitajima & Yamashita, 1999; Mateeff et al., 1985; Soto-Faraco & Kingstone, in press), but apparently this immediate effect was not sufficient to produce a CAMA. One could also speculate that fixating the mobile visual stimulus caused eye movements that modulated the CAMA. However, given that in the stationary-sound condition, as in the other conditions, participants were fixating the mobile visual stimulus throughout the induction phase, this seems highly unlikely.

Rather, the present data are compatible with the notion that there is a direct perceptual link between the visual and auditory motion-processing systems. In this view, visual motion affects auditory motion processing, but visual motion does not affect motion processing of a stationary sound, as a stationary sound does not activate the auditory motion-processing system. At present, there is indeed some evidence that the neural processing of auditory motion involves neural mechanisms distinct from those involved in the processing of stationary

Table 1. *Aftereffects calculated from the estimated stationary stimulus (ESS) in the four conditions*

Condition	Induction schema: Rise-right/fall-left	Induction schema: Rise-left/fall-right	Average
Audiovisual congruent	0.56	0.75	0.66
Auditory only	0.22	0.15	0.19
Audiovisual incongruent	−0.46	−0.26	−0.36
Stationary sound	−0.06	0.08	0.01

sound location (Toronchuk, Stumpf, & Cynader, 1992). The present findings are also compatible with some recent functional magnetic resonance imaging data showing that during performance of both visual and auditory motion-discrimination tasks, certain areas of the brain are conjointly activated (Lewis, Beauchamp, & de Yoe, 2000).

REFERENCES

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

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

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

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

Dong, C., Swindale, N.V., & Cynader, M.S. (1999). A contingent aftereffect in the auditory system. *Nature Neuroscience*, *2*, 863–865.

Kitajima, N., & Yamashita, Y. (1999). Dynamic capture of sound motion by light stimuli moving in three dimensional space. *Perceptual and Motor Skills*, *89*, 1139–1159.

Lewis, J.W., Beauchamp, M.S., & de Yoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, *10*, 873–888.

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: Apparent auditory motion induced by a moving visual target. *Perception*, *14*, 721–727.

McCollough, C. (1965). Color adaptation of edge detectors in the human visual system. *Science*, *149*, 1115–1116.

Soto-Faraco, S., & Kingstone, A. (in press). Multisensory integration of dynamically-moving stimuli. In G. Calvert, C. Spence, & B.E. Stein (Eds.), *Handbook of multisensory processes*. Cambridge, MA: MIT Press.

Toronchuk, J.M., Stumpf, E., & Cynader, M.S. (1992). Auditory cortex neurons sensitive to correlates of auditory motion: Underlying mechanisms. *Experimental Brain Research*, *88*, 169–180.

Vroomen, J., Bertelson, P., & de Gelder, B. (1998). A visual influence in the discrimination of auditory location. In D. Burnham, J. Robert-Ribes, & E. Vatikiotos-Bateson (Eds.), *Proceedings of the International Conference on Auditory-Visual Speech Processing, AVSP'98* (pp. 131–135) [CD]. Adelaide, Australia: Causal Productions. (Available from AVSP'98, School of Psychology, UNSW, Sydney, NSW 2052, Australia)

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Auditory-visual spatial interactions: Automatic versus intentional components. In B. de Gelder, E. de Haan, & C. Heywood (Eds.), *Out of mind* (pp. 140–150). Oxford, England: Oxford University Press.

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

Vroomen, J., & de Gelder, B. (in press). Perceptual effects of cross-modal stimulation: The cases of ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence, & B.E. Stein (Eds.), *Handbook of multisensory processes*. Cambridge, MA: MIT Press.

Welch, R. (1999). Meaning, attention, and the "unity assumption" in 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.

Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in Cognitive Sciences*, *6*, 211–216.

(RECEIVED 5/20/02; REVISION ACCEPTED 8/31/02)