

Auditory and visual attention-based apparent motion share functional parallels

WENDY E. HUDDLESTON

University of Wisconsin, Milwaukee, Wisconsin

JAMES W. LEWIS

West Virginia University, Morgantown, West Virginia

RAYMOND E. PHINNEY, JR.

Wheaton College, Wheaton, Illinois

AND

EDGAR A. DEYOE

Medical College of Wisconsin, Milwaukee, Wisconsin

A perception of coherent motion can be obtained in an otherwise ambiguous or illusory visual display by directing one's attention to a feature and tracking it. We demonstrate an analogous auditory effect in two separate sets of experiments. The temporal dynamics associated with the attention-dependent auditory motion closely matched those previously reported for attention-based visual motion. Since attention-based motion mechanisms appear to exist in both modalities, we also tested for multimodal (audiovisual) attention-based motion, using stimuli composed of interleaved visual and auditory cues. Although subjects were able to track a trajectory using cues from both modalities, no one spontaneously perceived "multimodal motion" across both visual and auditory cues. Rather, they reported motion perception only within each modality, thereby revealing a spatiotemporal limit on putative cross-modal motion integration. Together, results from these experiments demonstrate the existence of attention-based motion in audition, extending current theories of attention-based mechanisms from visual to auditory systems.

Marquee lights on a theater sign can generate a strong, unambiguous percept of motion, from the timing of lights flashed successively at adjacent locations. On other marquees, however, alternate lights simply flash on and off in opposition. In the latter case, the potential motion cues are ambiguous: As one light goes off, both left and right neighbors come on, which can lead to motion perception in either direction. Such ambiguous displays can yield a single, coherent direction of perceived motion if the observer attentively tracks the successive onset of lights in one direction, yielding what has been termed "attention-based" motion (Cavanagh, 1992; Lu & Sperling, 1995, 2001; Verstraten, Cavanagh, & Labianca, 2000). Figure 1 illustrates another example of an ambiguous circular motion stimulus. When the two frames are alternately and repeatedly presented, an observer has the impression of an illusory square rotating in the center of the ring of disks. One can voluntarily switch the direction of apparent rotation by attentively following one of the corners of the illusory square in a desired direction, and this can occur even when the observer constantly fixates the center of the display and attends covertly to one of the corners.

Similar perceptual experiences can occur in other sensory modalities. For example, an observer can track a dog running along the opposite side of a solid fence from the location of its successive barks. However, it is not clear whether auditory attention can affect an ambiguous sound stimulus in a manner analogous to the effect of attention in vision. We hypothesized that an attention-based motion mechanism in the auditory domain does exist. Demonstrating such an attention-based auditory effect would identify an important high-level parallelism between the functional organization of vision and audition. The presence of an attention-dependent motion process in both modalities was necessary to demonstrate prior to raising the question of how attention might affect the processing of multimodal stimuli. Fujisaki and Nishida (2007) suggested that the perception of audiovisual synchrony might require such higher order processes as salient feature maps. We wanted to further test this hypothesis with multimodal motion perception.

Multiple models exist for demonstrating attention-based higher order motion processing in the visual system. We wanted to be unequivocal in the demonstration of such a

W. E. Huddleston, huddlest@uwm.edu

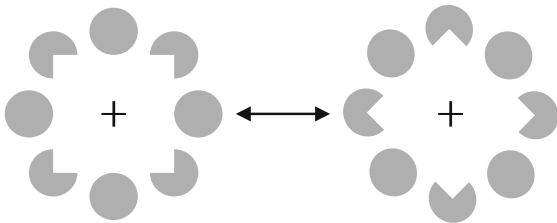


Figure 1. Ambiguous visual motion stimulus. When the two images are alternated in time, an observer perceives an “illusory” Kanizsa square rotating in the middle of the circular array. The direction of rotation can be voluntarily switched by the observer by a shift of attention.

mechanism in audition and, thus, created two different sets of experiments to address this issue. In the first two sets of experiments, we successfully demonstrate auditory attention-based motion perception and show that it has temporal properties comparable to an attention-mediated visual process. In a follow-up experiment, we demonstrate that subjects could track interleaved multisensory events only at low rotational speeds, but without having a coherent perception of motion across modalities. These results suggest that higher order motion mechanisms operate, at a minimum, in parallel within each modality. Furthermore, they extend previous concepts of high-order motion processing and clarify the role of attention and the computational level(s) at which attention can influence motion perception within and across modalities.

METHOD

We performed two different sets of experiments to test the phenomenon of auditory attention-based motion. Experiment 1 tested the ability of naive subjects to perceive coherent motion when listening to a purely ambiguous stimulus. In Experiment 2, subjects disambiguated an auditory stimulus by focusing attention to one of two target sounds. Lastly, we wanted to determine whether cues from one sensory modality could disambiguate an otherwise ambiguous stimulus presented in a different sensory modality. Subjects in Experiment 3 reported the motion direction of a stimulus with interleaved auditory and visual cues. The stimulus in this case was ambiguous within each modality but was unambiguous if subjects integrated the visual and auditory information.

Subjects

Two sets of naive subjects performed Experiment 1 ($n = 10$, 27.3 ± 5.2 years, 7 females) and Experiment 2 ($n = 9$, 29 ± 6.8 years, 3 females). Seven subjects completed Experiment 3 (32 ± 5.4 years, 2 females), 4 of whom had performed Experiment 2 as well. All subjects provided informed consent as approved by the Institutional Review Board of the Medical College of Wisconsin. All subjects reported normal or corrected-to-normal vision, normal hearing, and no neurological deficits. Subjects were screened for the ability to localize sound sources prior to testing.

Sound Stimuli

White noise for all experiments was independently generated for each speaker to optimize localizability in all experiments. For Experiment 2, two environmental sounds were selected for their ability to be localized in space and discriminated from one another. The first complex sound consisted of three temporal epochs containing four distinct bands of energy, having an overall higher perceived

pitch than did the second complex sound (see Figure 2). The power spectrum of the second complex sound showed energy widely distributed across a broad frequency range. These sounds were assessed psychophysically by 7 subjects to determine that the sounds were relatively devoid of more cognitive associations that might otherwise interfere with actual spatial perception in the paradigm.

The starting speaker position in Experiments 2 and 3 varied on each trial. Sequential tones at random speaker locations were added before and after the trial, to prevent the subject from determining the direction of the stimulus by simply tracking the order of the first or last two events. Each trial lasted 1–2 sec, with a total of one to four revolutions per trial. Forty-eight trials were completed at each frame duration. The order of conditions was counterbalanced across subjects. In Experiment 3, control stimuli were created in which only light-emitting diodes (LEDs) or only speakers were at each of the four locations. These control stimuli allowed for comparison of subject performance within modality with performance with the unambiguous multimodal stimuli.

Tasks

In Experiments 1 and 2, subjects reported their perceptions when exposed to a moving sound stimulus as described above. Observers sat in the center of a horizontal ring of speakers placed at ear height and judged the direction of motion of ambiguous and unambiguous sound stimuli as either “clockwise” (CW) or “counterclockwise” (CCW), if one imagined looking down on a horizontal clock face. Minimal instructions were given, to prevent any bias in perception for the subjects in Experiment 1. Subjects were instructed to attend to one of two distinct sounds for blocks of trials in Experiment 2. The order of attended sounds for each block was counterbalanced between subjects. In Experiment 3, subjects had to report the perceived direction of a visual, an auditory, or a multisensory stimulus rotation in a vertical array as CW or CCW. Further detail is given in the Results section.

Equipment

For Experiments 1 and 2, subjects sat in the center of a 6-ft-diameter ring of 8 Koss SA/30 speakers (Milwaukee, WI) positioned at the level of the ear and directed inward toward the ring’s center. Subjects were given a fixation point along the ring not at a location of a speaker and were told to maintain fixation throughout the experiments. In Experiment 3, subjects faced a vertical array containing LEDs on the vertical axis 10 in. from the fixation point and speakers on the horizontal axis 12 in. from the fixation point. Subjects sat between 29 and 35 in. from the array. They were positioned to optimize localization of the sound sources yet have the LEDs within their field of view. A chin bar stabilized the subject’s head. Each speaker and LED was independently controlled using Cool Edit Pro (Version 1.2, Adobe) or Cakewalk Pro Audio Version 9 (Cakewalk, Boston) with two Darla sound cards (Echo Digital Audio Corp., Carpinteria, CA). Speakers were volume adjusted by placing a precision integrating sound level meter (Brüel & Kjær, Nærum, Denmark) in the center of the array and manually adjusting it until all readings of a test sound were at the same volume (58 dBA).

Analysis

Data from Experiment 1 were qualitative and descriptive, so no formal statistical procedures were used. In Experiment 2, subjects attended to one of two complex sounds on each trial. Permutation tests (Good, 2000) were performed between the accuracy measures when subjects attended to the different sounds, to determine whether performance was significantly different between them. No significant difference in performance between the two sounds in the attention condition existed ($p > .05$); therefore, the data were collapsed across sound type for subsequent analyses. A two-way repeated measures ANOVA was performed with frame duration and task condition as the main factors. Main effects for both frame duration and task condition, as well as their interaction, were all

significant, so randomization tests for scores (Good, 2000) were performed to determine significant differences between task conditions at each frame duration. Five hundred twelve permutations were performed ($\alpha < .05$) for each test using NPFact software (Version 1.0, University of Victoria, Victoria, BC). A *t* test for comparison to a population mean was calculated for the unambiguous stimulus at 31, 62, and 125 msec and for the ambiguous stimulus at 31–250 msec to determine whether performance was significantly greater than chance ($\alpha = .05$).

RESULTS

Experiment 1: Motion Perception With Ambiguous Auditory Stimuli

Experiment 1 was designed to show that attention directed to an ambiguous auditory stimulus can lead to the perception of unambiguous auditory motion. Each of 10 naive subjects sat in the center of a horizontal, circular, eight-speaker array 6 ft in diameter (see Figure 3) and simply described what they heard when listening to a single 1-min trial while fixating a point between two speakers. Independently generated white noise was amplitude

modulated at 1 Hz, such that adjacent speakers were 180° out of phase. As with the visual stimulus illustrated in Figure 1, this auditory stimulus was completely ambiguous. Apparent motion in either a CW or CCW direction around the subject's head could be perceived by tracking the location of peak amplitude from speaker to speaker to perceive motion. All 10 subjects reported coherent rotational motion in a single direction for extended periods of time, and the direction could be changed voluntarily through shifts of attention.

A second set of trials was included in which each trial began with 1 sec of unambiguous motion in a randomly chosen direction, followed by 1 min of the ambiguous stimulus. The unambiguous motion was included to attempt to bias the perceived direction of motion of the subsequent ambiguous stimulus without using verbal instructions. Again, all 10 subjects reported coherent rotational motion. Eight subjects perceived motion in the same direction as the unambiguous stimulus, whereas 2 perceived an immediate shift in the perceived direction of the stimulus upon onset of the ambiguous component.

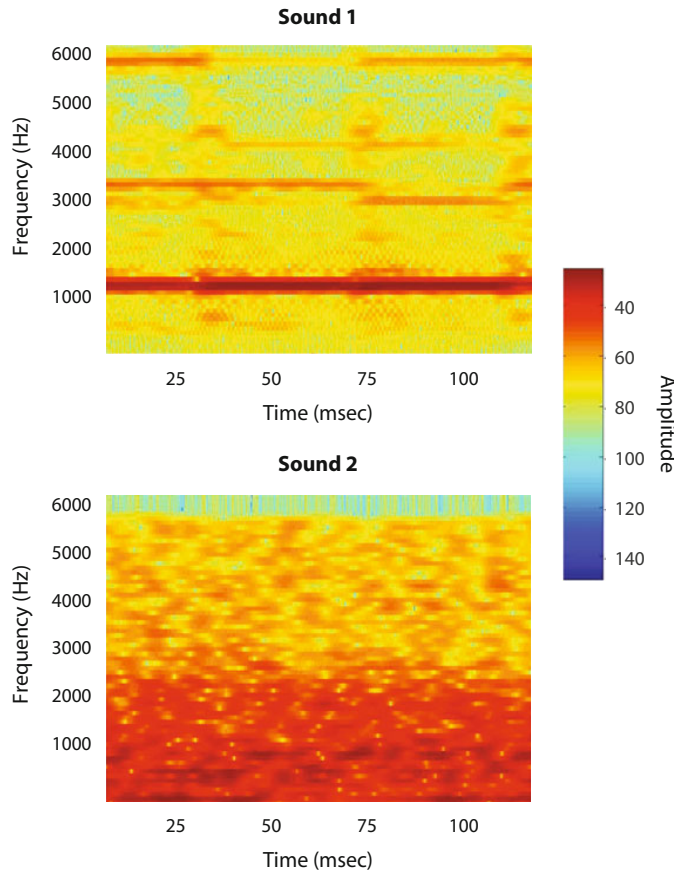


Figure 2. Spectrograms for complex sounds: Spectrograms for 125-msec samples of the two distinctive sounds used in Experiment 2. The scale for the amplitude of the signal at each frequency (*y*-axis) is represented to the right.

Extending this paradigm, subjects also listened to a 1.5-min version of the ambiguous stimulus and reported if and when the stimulus spontaneously changed directions. Verstraten and Ashida (2005) reported an average duration of motion perception in one direction of 1–3 sec before a spontaneous reversal of motion occurred when subjects attended to an ambiguous visual stimulus for extended periods of time. We anticipated that subjects would also have spontaneous reversals of direction with prolonged exposure to the ambiguous auditory stimulus. Indeed, this was the case with the 10 subjects who reported a range of zero to three reversals during the 1.5-min trial.

Finally, we tested the robustness of the perception of motion about the head at different rates of amplitude modulation between the two sets of speakers. Eight of the 10 subjects completed this portion of the experiment. Subjects listened to a 1-min sound clip and reported whether the sounds from the speakers seemed to be played simultaneously, seemed to be played in a random order, or created a strong perception of motion. Trials in which the sounds were amplitude modulated in a temporally congruent manner across all eight speakers (i.e., simultaneous amplitude modulation) were included occasionally as foils. The duration of the amplitude envelope (Figure 3) varied between 125, 250, 500, and 1,000 msec. At 125 msec, none of the 4 subjects who participated in that condition reported a perception of motion about the head but reported that the sounds were played either simultaneously or randomly. Conversely, at 500- and 1,000-msec durations, all 8 subjects reported a strong perception of motion. At the intermediate rate of 250 msec, 3 of 8 subjects reported good motion perception. These data suggest a robust perceptual experience

of attention-based motion about the head in which the direction of perceived motion could be changed at will and degraded with faster rates of amplitude modulation between alternate sets of speakers.

Experiment 2:

Disambiguation of an Auditory Stimulus

To quantify the characteristics of auditory attention-based motion observed in Experiment 1, we used a modified stimulus and paradigm (Figure 4). Subjects sat in the center of the same array, although only four speakers were used. Subjects performed two separate experimental conditions, both requiring them to identify the direction of perceived stimulus motion. First, subjects were asked to determine the direction of rotation (CW or CCW) of an unambiguous stimulus consisting of noise alternating with a distinctive sound presented sequentially around the speaker array (Figure 4A). Next, subjects determined the perceived direction of an ambiguous stimulus consisting of white noise and two distinct sounds. Subjects attended to one of the distinct marker sounds in a two-alternative forced choice design (Figure 4B). Attention directed to either marker, interleaved with white noise, should result in disambiguated motion. Figure 4B illustrates CW rotation when attending to Sound 1 and CCW rotation when attending to Sound 2. Both trial types contained all the same sounds, differing only in the cued attended marker, which was counterbalanced across blocks of trials. Subjects were instructed to attend to one of the two distinct sounds for all trials within one block. Verbal instructions were given prior to each block to specify the sound to be attended to. Each sound in the sequence was of equal duration, and this frame duration was varied systematically

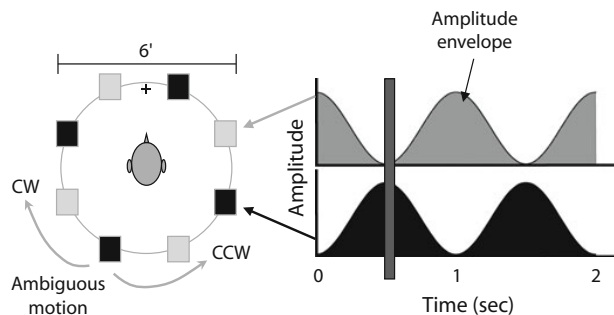


Figure 3. Experiment 1: Uncued attention-based motion perception. Subjects visually fixated a marker located between speakers while sitting in the middle of an eight-speaker array. Independently generated white noise was amplitude modulated in synchrony for four alternating speakers (gray boxes). The other four speakers (black boxes) were amplitude modulated 180° out of phase with the first four speakers. The two time series to the right illustrate the relationship of the amplitude modulation between sets of speakers. The vertical bar midway between Points 0 and 1 reflects a period of maximum amplitude in the latter speakers and minimum amplitude in the former speakers, as illustrated at the left of the figure. The inverted phase modulation between the two sets of speakers resulted in the illusion of an auditory object surrounding the head and body that could be perceived as rotating clockwise or counter-clockwise, and this ambiguous direction could be perceived to change through a voluntary shift in the allocation of attention by the observer.

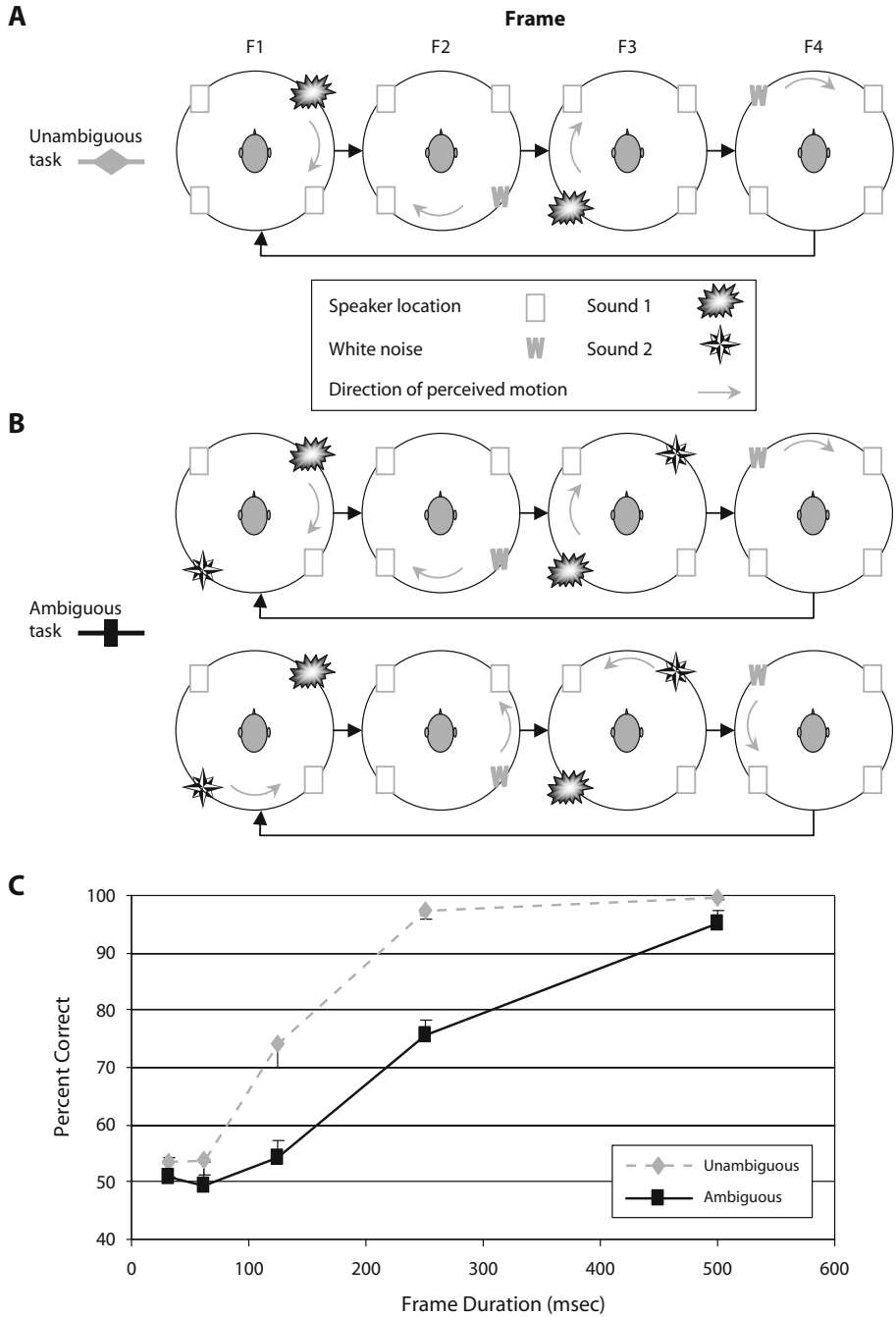


Figure 4. Experiment 2: Cued attention-based motion perception. Subjects visually fixated a marker located between speakers while sitting in the middle of a four-speaker horizontal array. (A) Unambiguous stimulus. Alternating white noise or distinctive sounds were played on a single speaker during each of four frames sequenced in a clockwise or counterclockwise direction (clockwise illustrated in figure). Subjects indicated perceived direction of motion on each trial. (B) Ambiguous stimulus. The stimulus is very similar to the unambiguous stimulus, but on frames when the first distinctive sound was played, a second (different) distinctive sound was played simultaneously at the opposite speaker. If the subject was cued to attend to the first sound, the perceived motion should rotate clockwise (direction of motion indicated by arrows between speaker locations). If the subject attended to the second distinctive sound in the same stimulus sequence, the perception should rotate in a counterclockwise direction. (C) Psychophysical performance for the unambiguous and ambiguous (requiring attention) tasks. Frame duration is the amount of time a sound was played at a particular location (interstimulus interval = 0). All values are means ± 1 SEM.

(31, 62, 125, 250, and 500 msec) to assess the temporal characteristics of the attention-based auditory motion percept.

Subject accuracy in determining the correct direction of rotation of the unambiguous (control) and ambiguous (attention) stimuli are graphed in Figure 4C. As the frame duration increased, the perceived rotational motion slowed and the correspondence between the perceived and predicted directions of motion (percentage correct) improved. A frame duration of 31 msec equals a rotation rate of 8.06 revolutions per second (rps), whereas a frame duration of 500 msec equals a rotation rate of 0.5 rps. Performance was not significantly different from chance for either the ambiguous or the unambiguous task at the shortest frame durations (Student's *t* test, $p > .1$ for 31 and 62 msec for both conditions). Subjects approached perfect performance in both tasks at the longest frame duration. However, the psychometric curve for the attention-based motion was shifted to the right at the intermediate frame durations; thus, it appears that attention-based auditory motion and unambiguous auditory motion may arise from temporally distinct mechanisms. Significant differences existed between the two task conditions at all frame durations longer than 62 msec [2 (task) \times 5 (frame duration) ANOVA: task \times frame duration interaction, $F(4,32) = 6.670$, $p < .0005$]. All subsequent randomization tests for differences in the scores (Good, 2000) of the two conditions were significant at $p < .05$.

Together, the data from Experiments 1 and 2 demonstrate the existence of an auditory attention-based motion mechanism that parallels the temporal dynamics of a similar mechanism in vision. To our knowledge, this is the first demonstration of this effect in the auditory domain. This might reflect a single "supramodal" mechanism operating for both sensory modalities. If so, this leads to the interesting prediction that attention-based motion might operate equally well for a stimulus composed of interleaved auditory and visual cues, which we sought to address in an additional experiment.

Experiment 3: Supramodal Apparent Motion

A supramodal attention-based motion mechanism might use a common representation, such as a salience feature map (Lu & Sperling, 1995), to encode features of a stimulus regardless of its modality (audio or visual). A salience feature map may transmit only information regarding salient stimulus location to a higher order motion processing mechanism and not necessarily information regarding the stimulus modality. We created a stimulus in which both audio and visual motion cues were individually ambiguous for circular motion, but if the spatial information of the cues were combined via a supramodal mechanism, a subject would perceive an unambiguous direction of motion. We tested this hypothesis by having participants attempt to fuse auditory and visual cues for a singular percept of circular motion.

For Experiment 3, we modified our stimulus array to be positioned in a vertical position and placed in front

of the subjects so they could view four LEDs and hear sounds coming from four speakers behind a cloth screen (the speakers were not visible). In a visual-only control condition, an LED was illuminated sequentially in one of the four spatial locations illustrated in Figure 5, which led to an unambiguous perception of motion in either a CW or CCW direction, similar to previously reported designs (Lakatos & Shepard, 1997). By analogy, for the auditory-only condition a sound was presented sequentially along the four locations, one location at a time, and subjects indicated the direction of rotation (CW or CCW). Participants had to bind consecutive events at different spatial locations for the perception of apparent motion. That is, participants had to perceive that a single object "moved" from one location to another in a sequential order to perceive circular motion (motion from top, right, bottom, and left locations for CW motion). For the multimodal condition (Figure 5), two speakers on the horizontal axis and two LEDs on the vertical axis were used. In this condition, the location of the sensory event rotated exactly as it did in the unimodal conditions, but now the sensory event would alternate between lights and sounds (top LED, right speaker, bottom LED, left speaker), with the path of sensory events following the same trajectory as in the unimodal conditions. The stimulus design was very similar to that in Experiment 2, but now the visual and auditory cues alternated (Figure 5B). The same range of frame durations was tested as in Experiment 2. If subjects could integrate across modalities, a single direction of rotation would be perceived. Note, however, that the cues within each individual modality were ambiguous for rotational motion. If subjects focused only on the lights, they would perceive vertical motion; conversely, if they focused only on the sounds, they would perceive horizontal motion. Seven subjects indicated the direction of perceived rotational motion when observing the multimodal stimulus. Surprisingly, none of the subjects had an integrated percept of rotational motion. Subjects could explicitly track the spatiotemporal order of the interleaved cues when the cues were sufficiently slow to infer the direction of rotation, but they did not subjectively perceive rotational motion per se. Rather, all subjects reported a percept of lights moving from one LED location to the other and of sounds moving from one speaker to the other independently in each modality, even though they were sequentially presented in alternate modalities over time.

Figure 6 illustrates the psychometric functions for performance on the direction discrimination task for unambiguous visual-only, auditory-only, and multimodal stimuli, all presented using the vertical array. As expected, subjects easily discriminated the direction of the visual stimulus, even at very fast frame rates, and subjectively reported the perception of a light rotating around the fixation point. Not unexpectedly, subjects had a more difficult time performing the auditory direction discrimination task on the vertical array than they did in Experiment 2, which was tested in the horizontal plane around the head. Even though subjects were tested for the ability to localize the two speakers on the horizontal axis,

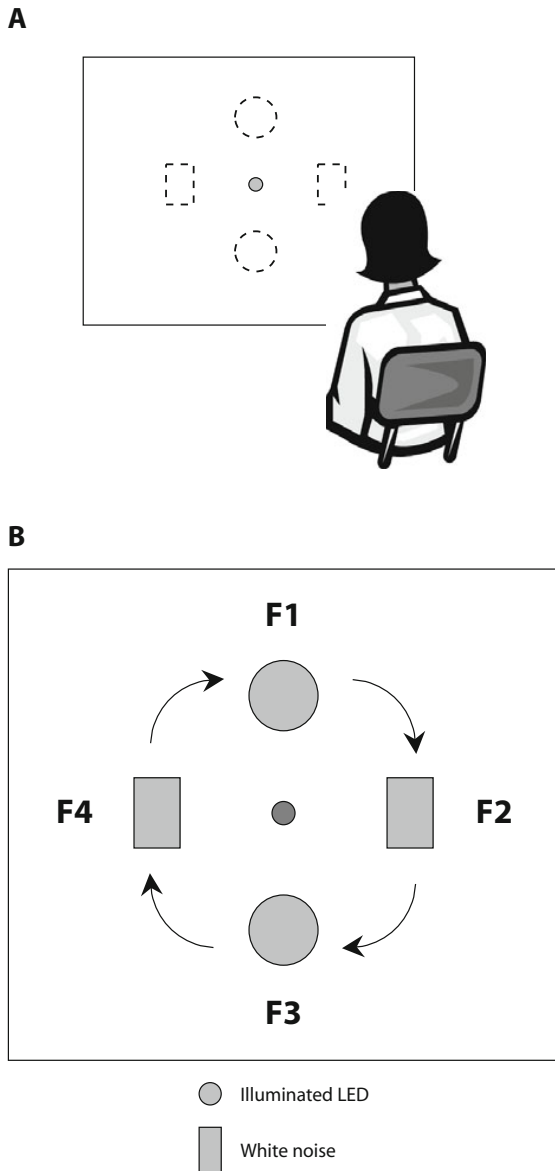


Figure 5. Experiment 3: Multimodal motion stimulus. (A) Vertical stimulus array consisting of LEDs (dashed circles) and speakers (dashed rectangles) was positioned behind black grille cloth so that LEDs could be seen when illuminated but speakers were not visible. Distance between the speakers was adjusted to ensure that each subject could localize the two speaker sounds along the horizontal axis. The LEDs were then positioned within the subject's field of vision as close to the distance of the speakers as possible. The angle between the head and the two speakers varied between 19.5° and 23.3° . The visual angle between the LEDs varied between 16.25° and 19.5° . The subject's head was stabilized with a chin bar (not shown). Subjects maintained central fixation throughout each experiment. (B) Illumination of a single LED (circle) alternated with white noise played from a single speaker (square) on successive frames (F1–F4). If the auditory and visual stimuli could be integrated sequentially, then the perception of motion should be in the clockwise direction. Subjects also performed control conditions in which the four locations were represented by all LEDs or all speakers.

it was difficult at the shorter frame rates to distinguish the speakers on the vertical axis, which led to poorer performance on the direction discrimination task. In the multimodal stimulus, removal of the auditory cues along the vertical axis essentially converted the auditory-only stimulus from being unambiguous to ambiguous. The multimodal stimulus could also be considered to be an ambiguous visual stimulus with the removal of the LEDs along the horizontal axis. Since the multimodal stimulus is ambiguous for rotational motion in both modalities, it would be expected that performance would decline in the multimodal condition in relation to the two unimodal unambiguous conditions.

Interestingly, intermediate levels of performance were achieved with the multimodal stimulus. Subjects could track the salient events across sensory modalities with the multimodal stimulus at a faster frame rate than with the auditory-only control condition. Performance with the multimodal stimulus was worse than it was with the visual-only condition, given the addition of the relatively poor localizability of the auditory stimulus. Conversely, performance was better with the multimodal stimulus than with the auditory-only stimulus, since the visual cues provided better localizability than did an auditory stimulus played in the same location.

DISCUSSION

In vision, directed spatial attention can lead to coherent motion perception in an otherwise ambiguous or illusory display (Cavanagh, 1992; Lu & Sperling, 1995, 2001; Verstraten et al., 2000). In Experiments 1 and 2, we demonstrate an analogous auditory effect for both implicitly and explicitly controlled attention. To our knowledge, this is the first evidence for a higher level, attention-based motion-processing mechanism in the auditory system. Additionally, we showed that the auditory motion percept for an ambiguous display suffers if the stimulus cues are presented too rapidly, consistent with attention-mediated mechanisms described in the visual system (Verstraten et al., 2000). Thus, it appears that attention may operate in an analogous fashion in both vision and audition, at least with respect to ambiguous motion displays. These results demonstrate an important parallel in motion processing between the two modalities.

Two basic models for the role of attention in higher level motion processing have been previously described (Cavanagh, 1992; Lu & Sperling, 1995, 2001; Verstraten et al., 2000), and we designed our auditory stimuli to test these two seemingly different models. The first model, attentive tracking-based motion perception (Cavanagh, 1992; Verstraten et al., 2000), proposes that the shift in attention is the basis for the motion calculation. The stimuli used to describe this model have objects in opposite frames equally spaced from one another, leading to ambiguity in the direction of motion perception in the absence of directed attention. The stimuli from Experiment 1 most closely match these criteria. In audition, as seen in vision, disambiguation of directional motion perception with these stimuli did not require training. Naive

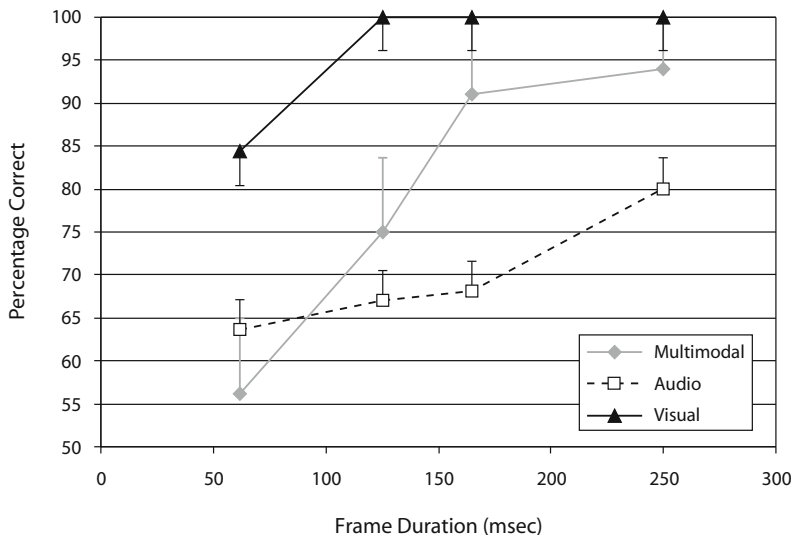


Figure 6. Direction discrimination performance with single modality and multimodal stimuli. Accuracy is plotted along the y-axis for the three experimental conditions: Multimodal stimulus (lights and sounds), auditory only, and visual only. Frame duration along the x-axis is the amount of time a sound was played or a light illuminated at a particular location. All values are means ± 1 SEM.

subjects in Experiment 1 perceived circular motion, so the results of the present study closely match those found for the visual system, using this type of experimental paradigm.

The second model, the feature salience model (Lu & Sperling, 1995, 2001), proposes that salient cues within the visual modality are plotted onto a common salience map, regardless of the specific feature of the salient cue. From this map, higher level processes such as motion direction are calculated. Motion stimuli used to describe the feature salience model remove or abstract any overall motion visible to first- and second-order visual systems. Another unique feature is the addition of alternate frames to an ambiguous stimulus. These alternate frames contain multiple features, which, by attending to a particular feature, subjects perceive as disambiguated motion. Differently from viewing a completely ambiguous stimulus, in which the perceived direction of movement is left to the discretion of the observer, the addition of alternate frames disambiguates the direction of motion when a particular feature in those frames is attended to; that is, opposite motion directions can be perceived, depending on which feature is attended to in the same stimulus. In the feature salience model, attention-based motion can be seen with very brief stimulus presentations too quick for attentive tracking, but requiring extensive training for motion perception to occur. Our stimuli for Experiment 2 more clearly test for an auditory analogue to the feature salience model.

As with viewing visual stimuli designed in this manner, subjects reported motion perception in the opposite direction when instructed to attend to one versus the other

sound in Experiment 2. Interestingly, subjects performed this task without intensive training and were able to attend to the other sound cue without additional training during data collection, unlike the results reported by Lu and Sperling (1995). This apparent discrepancy in training may have been due to relatively longer presentations of the stimuli at each respective location in the present experiments. Once again, the results of the present study closely match those found for the visual system using this second type of experimental paradigm. The stimuli designed for the present set of auditory experiments satisfied the cue feature criteria for both models; thus, the results are suggestive of a true auditory analogue of visual attention-based motion in audition.

When comparing our results to those in vision, the temporal dynamics of attention-based motion in vision and audition appear to be similar. Verstraten et al. (2000) reported the temporal limit for attentional tracking of an ambiguous visual stimulus to be 1 rps, which fits nicely with our present data from the auditory system. The speeds at which subjects could successfully determine motion direction are slightly slower than those reported by Lu and Sperling (1995), who, however, employed extensive training for their subjects, which was not done in the present set of experiments. As in vision, attention-based auditory motion appears to be more sluggish than purely stimulus-driven motion does (Experiment 2), as suggested by the degraded performance at 1–2 rps but not at 0.5 rps.

The similar temporal properties of attention-based motion in vision and audition might be consistent with a common computational mechanism (Farah, Wong,

Monheit, & Morrow, 1989; Shimojo & Shams, 2001; Ward, McDonald, & Lin, 2000). Consistent with this hypothesis, reports of cross-modal effects for both spatial (Eimer, van Velzen, & Driver, 2004; Frassinetti, Bolognini, & Ládavas, 2002; Giard & Peronnet, 1999; Macaluso, Frith, & Driver, 2002; McDonald, Teder-Sälejärvi, & Hillyard, 2000; Spence & Driver, 1996) and nonspatial (Driver, 1996; Fort, Delpuech, Pernier, & Giard, 2002; King & Calvert, 2001; Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002) attention tasks are many. Also, interactions between auditory and visual motion processing suggest a common motion mechanism (Berman & Colby, 2002; Bremmer et al., 2001; Driver & Spence, 2000; Kitagawa & Ichihara, 2002; Lakatos, 1995; Lakatos & Shepard, 1997; Lewis, Beauchamp, & DeYoe, 2000; Soto-Faraco, Kingstone, & Spence, 2003). Typically, spatial and temporal congruence of stimuli in the two modalities is required for subjects to integrate visual and auditory cues into a coherent motion percept (Meyer, Wuerger, Röhrbein, & Zetsche, 2005). However, in Experiment 3, a particular motion path could be specified *only* by combining cues, which were both temporally and spatially incongruent, across modalities. The results of Experiment 3 provide initial evidence that, although subjects could use information from both modalities to determine the trajectory of the stimulus, the stimulus used in this experiment was not sufficient to overcome the need for spatial and temporal congruence to integrate multimodal cues for the *perception of motion* across modalities and, therefore, did not lead to the perception of a unified “audiovisual” stimulus. It remains to be determined whether other temporal and/or spatial constraints are necessary to “bind” the audio and visual motion inputs to form a coherent percept of a “multimodal” object.

In sum, the results of this study demonstrate auditory attention-based motion perception, thereby extending our previous study identifying a second-order auditory motion percept (Huddleston & DeYoe, 2003). We developed stimuli based on two models of visual attention-based motion to illustrate a parallelism between visual attention-based motion mechanisms and auditory attention-based motion mechanisms. We also illustrated the temporal dynamics of attentional tracking across sensory events that are not temporally or spatially congruent. Together, these studies show a functional parallelism between visual and auditory motion perception that extends beyond the first-order level of luminance/amplitude-based processing. These results also extend previous concepts of high order motion processing and clarify the role and computational level(s) at which attention can influence motion perception within and across modalities.

AUTHOR NOTE

This work was supported by Grants EY10244, MH51358, and RR0058. We thank T. Prieto and T. Thelaner for their assistance with hardware design. Correspondence concerning this article should be addressed to W. E. Huddleston, Department of Human Movement Sciences, University of Wisconsin, P. O. Box 413, Milwaukee, WI 53201-0413 (e-mail: huddlest@uwm.edu).

REFERENCES

- BERMAN, R. A., & COLBY, C. L. (2002). Auditory and visual attention modulate motion processing in area MT+. *Cognitive Brain Research*, *14*, 64-74.
- BREMMER, F., SCHLACK, A., SHAH, N. J., ZAFIRIS, O., KUBISCHIK, M., HOFFMANN, K., ET AL. (2001). Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, *29*, 287-296.
- CAVANAGH, P. (1992). Attention-based motion perception. *Science*, *257*, 1563-1565.
- 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. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*, R731-R735.
- EIMER, M., VAN VELZEN, J., & DRIVER, J. (2004). ERP evidence for cross-modal audiovisual effects of endogenous spatial attention within hemifields. *Journal of Cognitive Neuroscience*, *16*, 272-288.
- FARAH, M. J., WONG, A. B., MONHEIT, M. A., & MORROW, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia*, *27*, 461-470.
- FORT, A., DELPUECH, C., PERNIER, J., & GIARD, M.-H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audiovisual object detection in humans. *Cerebral Cortex*, *12*, 1031-1039.
- FRASSINETTI, F., BOLOGNINI, N., & LÁDAVAS, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*, 332-343.
- FUJISAKI, W., & NISHIDA, S. (2007). Feature-based processing of audiovisual synchrony perception revealed by random pulse trains. *Vision Research*, *47*, 1075-1093.
- GIARD, M. H., & PERONNET, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*, 473-490.
- GOOD, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses* (2nd ed.). New York: Springer.
- HUDDLESTON, W. E., & DEYOE, E. A. (2003). First-order and second-order spectral ‘motion’ mechanisms in the human auditory system. *Perception*, *32*, 1141-1149.
- KING, A. J., & CALVERT, G. A. (2001). Multisensory integration: Perceptual grouping by eye and ear. *Current Biology*, *11*, R322-R325.
- KITAGAWA, N., & ICHIHARA, S. (2002). Hearing visual motion in depth. *Nature*, *416*, 172-174.
- LAKATOS, S. (1995). The influence of visual cues on the localisation of circular auditory motion. *Perception*, *24*, 457-465.
- LAKATOS, S., & SHEPARD, R. N. (1997). Constraints common to apparent motion in visual, tactile, and auditory space. *Journal of Experimental Psychology: Human Perception & Performance*, *23*, 1050-1060.
- LEWIS, J. W., BEAUCHAMP, M. S., & DEYOE, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, *10*, 873-888.
- LU, Z.-L., & SPERLING, G. (1995). Attention-generated apparent motion. *Nature*, *377*, 237-239.
- LU, Z.-L., & SPERLING, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America A*, *18*, 2331-2370.
- MACALUSO, E., FRITH, C. D., & DRIVER, J. (2002). Directing attention to locations and to sensory modalities: Multiple levels of selective processing revealed with PET. *Cerebral Cortex*, *12*, 357-368.
- MCDONALD, J. J., TEDER-SÄLEJÄRVI, W. A., & HILLYARD, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*, 906-908.
- MEYER, G. F., WUERGER, S. M., RÖHRBEIN, F., & ZETZSCHE, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, *166*, 538-547.
- SHIMOJO, S., & SHAMS, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology*, *11*, 505-509.
- SOTO-FARACO, S., KINGSTONE, A., & SPENCE, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, *41*, 1847-1862.

- SOTO-FARACO, S., LYONS, J., GAZZANIGA, M., SPENCE, C., & KINGSTONE, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, **14**, 139-146.
- SPENCE, C., & DRIVER, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1005-1030.
- VERSTRATEN, F. A. J., & ASHIDA, H. (2005). Attention-based motion perception and motion adaptation: What does attention contribute? *Vision Research*, **45**, 1313-1319.
- VERSTRATEN, F. A. J., CAVANAGH, P., & LABIANCA, A. T. (2000). Limits of attentive tracking reveal temporal properties of attention. *Vision Research*, **40**, 3651-3664.
- WARD, L. M., McDONALD, J. J., & LIN, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, **62**, 1258-1264.

(Manuscript received May 25, 2007;
revision accepted for publication March 24, 2008.)