

Research Article

In the Footsteps of Biological Motion and Multisensory Perception

Judgments of Audiovisual Temporal Relations Are Enhanced for Upright Walkers

Ayse Pinar Saygin,^{1,2} Jon Driver,¹ and Virginia R. de Sa²

¹Institute of Cognitive Neuroscience, University College London, and ²Department of Cognitive Science, University of California, San Diego

ABSTRACT—*Observers judged whether a periodically moving visual display (point-light walker) had the same temporal frequency as a series of auditory beeps that in some cases coincided with the apparent footsteps of the walker. Performance in this multisensory judgment was consistently better for upright point-light walkers than for inverted point-light walkers or scrambled control stimuli, even though the temporal information was the same in the three types of stimuli. The advantage with upright walkers disappeared when the visual “footsteps” were not phase-locked with the auditory events (and instead offset by 50% of the gait cycle). This finding indicates there was some specificity to the naturally experienced multisensory relation, and that temporal perception was not simply better for upright walkers per se. These experiments indicate that the gestalt of visual stimuli can substantially affect multisensory judgments, even in the context of a temporal task (for which audition is often considered dominant). This effect appears to be constrained by the ecological validity of the particular pairings.*

Natural perception is inherently multisensory, involving the processing and integration of information from multiple modalities. In the auditory and visual modalities alone, there are many examples of frequently encountered stimuli that are more

often bimodal than unimodal (e.g., moving objects that create predictable noises, speech).

Although the majority of laboratory experiments on perception focus on one sensory modality at a time, multisensory perception has been an area of interest to psychologists for several decades, and research in this area has expanded in recent years (Calvert, Spence, & Stein, 2004; Spence & Driver, 2004). In particular, there is now a growing literature on perception of temporal relations between visual and auditory events (e.g., Alais & Carlile, 2005; Guttman, Gilroy, & Blake, 2005; Shams, Kamitani, & Shimojo, 2002; Vroomen & de Gelder, 2000; Zampini, Shore, & Spence, 2005). But most of this research (with the notable exception of studies inspired by the McGurk effect in the speech domain; McGurk & MacDonald, 1976) has used relatively impoverished, simple visual and auditory events; therefore, little is known about factors affecting the perception of multisensory temporal relations in more complex, meaningful or ecological stimuli.

In the experiments reported here, we studied perception of audiovisual temporal relations in a class of stimuli that are relatively natural and tap into extensive ecological experience, yet are also highly controllable: stimuli portraying *biological motion*. Ever since Johansson (1973) first reported that the motion trajectories of a few points on the human body could suffice for perception of various human actions, there has been an extensive literature on perception of biological motion from point-light displays (see Blake & Shiffrar, 2007, for a review). Most studies of biological-motion perception, with just a few exceptions (e.g., Arrighi, Alais, & Burr, 2006; Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Brooks et al., 2007), have

Address correspondence to A.P. Saygin, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, WC1N 3AR, London, United Kingdom, e-mail: a.saygin@fil.ion.ucl.ac.uk.

utilized unisensory visual stimuli only. However, in natural settings, perception of biological movements is often multisensory, as visual information is accompanied by related inputs in other modalities, notably, audition (as when footsteps are heard as well as seen). In this study, we considered possible implications of the gestalt of biological motion for perceiving audiovisual temporal relations between seen point-light stimuli and auditory events.

It is well established that point-light walkers are less recognizable and appear less coherent when they are inverted, rather than upright (e.g., Pavlova & Sokolov, 2000; Shipley, 2003; Sumi, 1984; Tadin, Lappin, Blake, & Grossman, 2002). But note that inversion does not alter the temporal information contained within the local visual motions. We examined whether judgments of the relation between auditory timing and visual timing might nevertheless be more accurate when the judgments concern upright, rather than inverted, point-light walkers (Experiment 1). Establishing that this was the case provided initial evidence that the gestalt of the upright walker has cross-modal consequences for audiovisual temporal judgments. Next, we sought to determine whether upright point-light walkers showed a similar advantage over scrambled point-light walkers, which contain the same motions as upright walkers, but without the gestalt of a walking figure (Experiment 2). Finally, we investigated whether the advantage for upright walkers still holds in a less ecologically valid situation: Specifically, we offset the phases of the visual and auditory stimuli, so that the sounds could not easily be heard as footsteps (Experiment 3).

GENERAL METHOD

Participants were adults ages 18 through 34. They reported normal or corrected-to-normal visual acuity and normal hearing, and had no known psychiatric, neurological, or cognitive abnormalities. Each participant gave informed consent in accord with local ethics.

On each trial, participants viewed periodically moving white dots presented on a uniform black background. At the same time, we also presented periodic sounds (sequences of beeps, discussed later in this section). The task was to indicate whether

the auditory and visual stimuli had matching or mismatching temporal frequencies.

To produce the visual stimuli, we implemented Cutting's (1978) classic algorithm for generating point-light walkers using Matlab (Natick, MA). These stimuli have the virtue of giving rise to vivid perception of a person walking, while being highly controllable and periodic. The figures were defined by 11 point lights, some of which were briefly occluded during the motion trajectory (e.g., the elbow dot could disappear behind the "torso"). The height of the point-light figures subtended approximately 5.5° of visual angle when the figures were viewed from 60 cm. The direction in which the point-light animations faced, right or left, was determined randomly on each trial. The figures did not translate on the screen when "walking," but remained at the center of the display (Fig. 1).

The auditory stimuli were sequences of 1,000-Hz beeps; the duration of each beep was 100 ms. The beeps were presented binaurally over headphones, at rates that varied across trials.

Participants were told that on each trial, they would see about a dozen periodically moving white dots on a black background, and that at around the same time, they would hear periodically presented beeps. They were told that the movement and the sounds would each have a constant periodicity and were asked to judge whether or not the movement and sounds had the same temporal frequency. They indicated by button press whether there was a temporal match or mismatch between the visual and auditory cycles, regardless of the kind of motion they saw on the screen. The experiment began with 12 practice trials; by the end of these trials, all participants indicated that they understood the task.

Each trial started with a fixation point displayed for 500 ms. For the next 2,000 ms, a point-light animation was presented at 60 frames/s, together with a sequence of sounds. After the stimulus presentation, participants pressed one of two keys to indicate either a match or a mismatch between the auditory and visual temporal frequencies. After each response, a visual feedback cue (green dot = correct; red dot = incorrect) appeared for 500 ms. Testing sessions had multiple blocks comprising 30 trials each. Each of several difficulty levels (defined by the actual temporal-frequency mismatch between the visual and auditory stimuli, as described later in this section) was tested in two

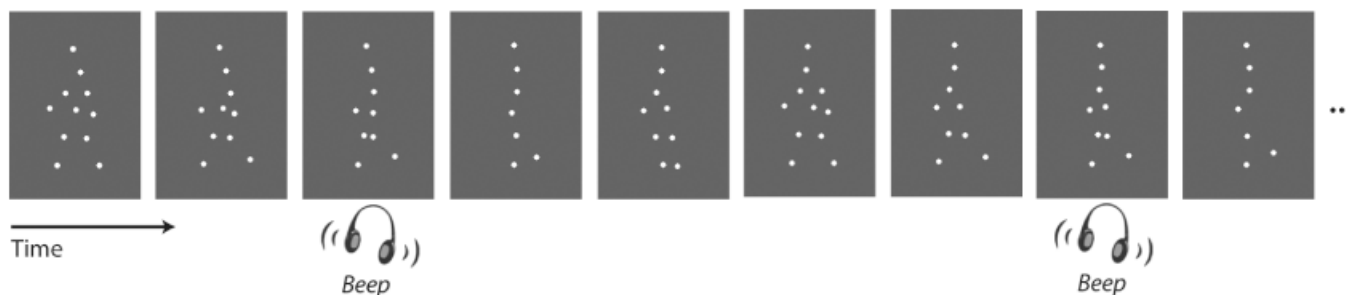


Fig. 1. Schematic of the visual and auditory sequences. In this example, the frames show an upright point-light walker. The example is taken from a match trial, so the sounds are simultaneous with frames in which one or the other foot dot appears to hit the ground and change movement trajectory.

separate blocks, for a total of 60 trials per level. The order of blocks was pseudorandomized (the first blocks at all levels were administered before the second blocks) and counterbalanced across participants. Each testing session lasted 70 to 90 min depending on the duration of the breaks the participant took between blocks.

The average temporal frequency of the visual stimuli was 2 Hz, with one cycle corresponding to one footstep. Thus, it took on average 1 s for two footsteps (one with each foot) to be completed. The 2-Hz rate was an average value because the frequency (and thus the walking speed) was jittered randomly (and continuously) between trials by up to ± 0.2 Hz. The visual percepts at these frequencies tend to correspond to natural, biologically plausible walking speeds (Beintema, Oleksiak, & van Wezel, 2006). Because each trial was 2 s long, the frequency jitter introduced into the visual stimuli meant that some trials contained slightly more (or fewer) cycles than others, but, on average, there were four footsteps per trial (two steps with each foot).

Half the trials in each block were randomly selected to be *match trials*, and the other half were *mismatch trials*. On match trials, the auditory stimuli consisted of tones synchronized with the walker's footsteps such that a beep sounded each time a foot dot made apparent contact with an apparent surface to reverse direction. Thus, in this case, after the walker's frequency was selected, the same frequency was selected for the auditory stimuli. On mismatch trials, the frequencies of the auditory and visual stimuli were not the same. The frequency for the auditory stimuli was determined by subtracting a value between 0.01 and 0.4 Hz from the walker's frequency (in Experiment 3, a value between 0.02 and 0.8 Hz was subtracted). For example, in a mismatch trial with a 2.11-Hz walker and a frequency offset of 0.08 Hz, the sound-repetition frequency would be 2.03 (2.11 minus 0.08). The smaller the offset value, the more difficult it was to detect a mismatch. Hence, the frequency offset between the visual and auditory stimuli in mismatch trials constituted the difficulty level. There were seven blocked difficulty levels in Experiments 1 and 2 (0.01, 0.02, 0.05, 0.08, 0.12, 0.2, and 0.4); Experiment 3 had six, slightly different, difficulty levels (0.02, 0.04, 0.08, 0.12, 0.4, and 0.8). On mismatch trials, the sounds always had a lower repetition frequency than the walkers. The difference in frequency led to increasing asynchronies as the trial unfolded; however, this increasing offset never realigned the sounds with footsteps at a later point, given the relatively short trial length.

Each visual footstep was defined as the frame in the animation corresponding to when a single foot dot changed trajectory, as if hitting an apparent surface. This point occurred 10% into the movement corresponding to a step. For example, if the walker took exactly 1 s to complete two steps (i.e., at 2 Hz), the footsteps would occur at 0.1 and 0.6 s.

In Experiments 1 and 2, when the temporal frequencies of the visual and auditory stimuli matched, the sounds corresponded to

the footsteps of the walkers. On both match and mismatch trials, the first sound always coincided with the first visual footstep. In Experiment 3, we deliberately offset the phase of the sounds away from the phase of the footsteps (by half of the step cycles), even when the frequency matched.

Stimuli were presented and responses were collected using the Psychophysics Toolbox for Matlab (Brainard, 1997). We used bootstrapping methods to estimate each participant's perceptual thresholds and to compare conditions within participants and across groups of participants. Psychometric functions were fit using the *psignifit* toolbox for Matlab, which uses a maximum likelihood method (Wichmann & Hill, 2001a). Statistical significance of the differences between thresholds was calculated using similar methods, and using the same toolbox plus the *pcmp* toolbox (Wichmann & Hill, 2001b). We also used standard repeated measures analyses of variance (plus, for completeness, Mann-Whitney *U* tests to circumvent the assumptions of parametric statistics) when comparing effects between experiments. Finally, the data were also examined within the signal detection framework. We report only accuracy data because the results of the sensitivity analyses paralleled those of the accuracy analyses and no further insights emerged.

EXPERIMENT 1

In Experiment 1, we contrasted upright and inverted point-light walkers. Half the trials showed upright walkers, and the other half showed inverted walkers; the order of the trials was random. The same walker was used on all trials; whether the walker faced to the right or to the left (mirror image about the vertical) was randomly determined for each trial. Eight adults (5 females, 3 males) participated and performed a two-alternative forced-choice task (same or different repetition frequencies) on each trial.

The critical result was that participants' judgments were more accurate when the walkers were upright than when they were inverted, even though these two kinds of trials contained identical temporal-frequency information. Figure 2a shows the combined data from all participants as an overall psychometric function. The figure shows that performance was better in the upright-walker condition. The difference between these two psychophysical curves was significant: Using bootstrapping (see General Method), we found that the likelihood that the curves came from the same underlying distribution was very small ($p = .001$). Furthermore, for each of the 8 participants, the estimated 75%-accuracy threshold was significantly lower for upright walkers than for inverted walkers ($p < .05$, two-tailed), as established via bootstrapping analyses on each participant's data.

EXPERIMENT 2

In Experiment 2, we sought to generalize the finding from Experiment 1 by comparing participants' accuracy for upright

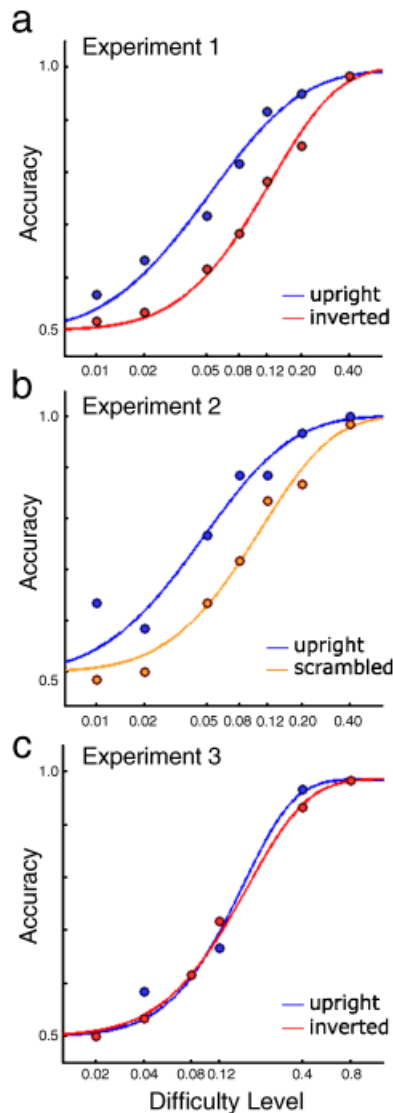


Fig. 2. Psychophysical curves depicting data from all subjects in (a) Experiment 1, (b) Experiment 2, and (c) Experiment 3. Accuracy (proportion correct) is plotted as a function of the difficulty level of a block, that is, the frequency difference (in hertz, plotted on a log scale) between the sound sequence and the visual motion in the mismatch trials within that block; this difference corresponds to task difficulty (see General Method).

versus scrambled walkers. Across a range of methodologies, biological-motion research has used scrambled stimuli extensively as control stimuli (e.g., Grossman & Blake, 2002; Ikeda, Blake, & Watanabe, 2005; Pavlova, Lutzenberger, Sokolov, & Birbaumer, 2004; Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004).

An upright walker was the visual stimulus in half the trials, and the remaining half presented a scrambled walker; the two kinds of stimuli were intermingled randomly. Our new control visual stimuli were created by spatial scrambling, that is, by randomizing the starting positions of the dots while keeping their local motion trajectories intact. The scrambled walkers thus had the same (local) information about temporal frequency as the

intact walkers, but the relations among visual dots (and therefore the gestalt) were altered. A single scrambled animation and its mirror image (about the vertical axis) were used throughout the experiment; similarly, a single upright walker and its mirror image (about the vertical) were used as the right- and left-facing upright walkers. As before, the first sound in the sequence always coincided with the change in trajectory for one or the other foot dot, although in the case of the scrambled walkers, these foot dots appeared in scrambled locations with respect to each other and to the other dots.

One participant from Experiment 1 and 3 new participants were tested in Experiment 2 (2 females, 2 males). Their performance was significantly better with the upright walkers than with the scrambled walkers. The overall psychophysical curves for the upright and scrambled stimuli, shown in Figure 2b, differed significantly from each other ($p < .001$). For 3 of the 4 participants, the estimated 75%-accuracy threshold was significantly lower for upright walkers than for scrambled walkers ($p < .05$, two-tailed); the 4th participant showed a trend in the same direction, but it did not reach significance.

EXPERIMENT 3

Our final experiment tested whether the effect of upright walkers observed in Experiments 1 and 2 would persist or disappear when sounds with matching temporal frequency no longer coincided with the plausibly sound-producing visual event of the footstep. Instead of being aligned with a visual footstep, the first sound on each trial occurred 50% into the first footstep, and thus never coincided with a visual footstep or with change in the trajectory of another dot. Because this task was harder overall, we used slightly different difficulty levels (corresponding to frequency offsets of 0.02, 0.04, 0.08, 0.12, 0.4, and 0.8) in Experiment 3, in order to obtain reliable psychophysical measurements. The task and procedure were otherwise the same as in Experiment 1.

If the observed advantage in perception of temporal frequency when walkers were intact and upright reflected some nonspecific general advantage within vision alone, then participants in Experiment 3 also would have been expected to perform better when judging upright walkers than when judging inverted walkers. If instead the previously observed advantage in judging intact upright walkers reflected the (natural) temporal correspondence between the gestalt of the visual walker and the auditory footsteps, then this advantage would have been expected to diminish or disappear in Experiment 3.

One participant from Experiment 1 and 3 new participants were tested (2 females, 2 males). The results were qualitatively different from those of Experiment 1, showing no difference between upright and inverted walkers in this new, phase-offset situation. The psychophysical curves, shown in Figure 2c, did not differ from each other; the bootstrapping analysis revealed that these data likely came from the same distribution (even a

one-tailed test indicated no difference, $p = .4$). Moreover, there were no significant differences between the two conditions in any of the individual participants' estimated 75%-accuracy thresholds (even with one-tailed tests, all $ps > .1$).

To confirm the difference in outcome for Experiments 1 and 3 more formally, we calculated the difference between the 75%-accuracy thresholds for inverted and upright walkers (inverted minus upright) in each experiment ($n = 8$ in Experiment 1, $n = 4$ in Experiment 3). The inversion effect in Experiment 1 was indeed significantly larger than the (null) effect in Experiment 3, as confirmed both with a between-subjects F test, $F(1, 11) = 12.25$, $p = .003$, one-tailed, and with a nonparametric Mann-Whitney U test, $z = -2.717$, $p = .003$, one-tailed. Thus, the advantage for intact upright walkers was specific to the situation in which seen walking and heard footsteps corresponded naturally, and was not a nonspecific advantage for upright walkers in general.

DISCUSSION

In a series of experiments, we found clear evidence that multisensory judgments comparing the temporal frequency of auditory and visual cycles were performed better when matching auditory and visual events corresponded to the footsteps of an upright point-light walker. When the gestalt of the walker was disrupted (by inversion in Experiment 1, by scrambling in Experiment 2), the audiovisual comparison was impaired. The advantage for upright over inverted walkers disappeared when the sounds were no longer phase-locked to the footsteps of the visual walker (Experiment 3). This implies that the benefit we observed when the stimuli were phase-locked does not simply reflect better perception for upright visual walkers per se, but rather reflects better perception specifically when sounds are synchronized with particular visual events that could plausibly produce them (e.g., heard footsteps).

A previous purely visual study (Tadin et al., 2002) suggested that upright point-light walkers can provide a beneficial intrinsic "reference frame" that may allow more efficient encoding of local features. But the dependence of the present cross-modal effect on the particular phase relation between auditory and visual footsteps (Experiments 1 vs. 3) indicates that the benefit for intact upright walkers we observed is not purely visual, but instead reflects the natural temporal relation between the visual cycle and a corresponding auditory cycle that can appear to be caused by the related visual events.

Recent experiments (Guttman et al., 2005) with very different stimuli showed that visual rhythms may be automatically encoded in the auditory domain, which arguably may specialize in processing temporal structure (Welch, 1999). But note that in the present experiments, the visual gestalt of the walker, and not just the temporal structure, played a critical role (and the temporal structure was the same for the upright walkers as for the inverted and scrambled walkers). An additional factor to consider is that perception of body movements, even in point-light

stimuli, can engage the viewer's own motor system (for neuroimaging and neuropsychological evidence, see Saygin, 2007, and Saygin et al., 2004); thus, temporal structure might also become encoded motorically when biological motion is perceived. Neuroimaging studies using our audiovisual paradigm may shed light on the brain systems affected, and whether they include regions conventionally linked to biological motion (e.g., the posterior superior temporal sulcus; Barraclough et al., 2005; Grossman & Blake, 2002; Saygin et al., 2004), modality-specific auditory and visual areas (e.g., Noesselt et al., in press), or motor components of the mirror system (Rizzolatti & Craighero, 2004).

We have suggested that our results reflect an advantage in cross-modal perception of temporal frequency when the visual gestalt of an intact upright walker bears a natural temporal relation to heard footsteps. Recently, Troje and Westhoff (2006) showed that foot dots convey special information in biological-motion displays and that they may be particularly important for inversion effects. One might wonder whether the entire walker is indeed necessary in our paradigm, or whether just the foot dots would suffice to obtain the same effect. In a control study with 4 new participants, we found that the advantage for upright over inverted presentation disappeared when only the two foot dots were shown. Further research with variants of our paradigm could establish exactly which aspects of the upright walker are necessary to produce our cross-modal effect.

Our results bring together traditionally separate psychological domains (perception of biological motion and multisensory integration) while suggesting further research directions. An association between seen and heard footsteps is often observed in daily life, although presumably with slight variations in audiovisual offset (e.g., slightly different auditory delays depending on viewing distance—see Arrighi et al., 2006; Spence & Squire, 2003). Future studies might examine whether or not perfect synchrony in the matching condition is optimal, as well as the extent to which observers can deal with slight fixed auditory delays or even adapt to larger delays over time, as has been demonstrated using different stimuli (Vroomen, Keetels, de Gelder, & Bertelson, 2004).

The fact that this study made use of an ecologically appropriate correspondence between visual and auditory footsteps also raises the question of whether observers become sensitive to such ecological pairings through experience. There is ample literature on infants' perception of audiovisual temporal correspondences (e.g., Lewkowicz, 2003; Spelke, Smith Born, & Chu, 1983), as well as a rich and growing literature on infants' perception of biological motion (e.g., Bertenthal, Proffitt, & Kramer, 1987; Bertenthal, Proffitt, Spetner, & Thomas, 1985; Reid, Hoehl, & Striano, 2006). But to our knowledge, there has been little or no developmental research bringing these two topics together, as we sought to join them in our study of adults. Adapting our paradigm for use with infants (e.g., by using selective looking or cross-modal habituation) might provide a way to test whether the cross-modal effect demonstrated in this study requires

hard-wired mechanisms for detecting audiovisual correspondence or instead requires experience with particular types of real-world stimuli in which such correspondence arises.

In this study, we sought to go beyond prior work on cross-modal temporal perception by using more complex, meaningful and ecological stimuli. But it should be acknowledged that our study is only an initial footstep (pun intended!) in this direction. To create our visual stimuli, we used Cutting's (1978) algorithm, which is well established in the field and has the virtue of allowing precise control of temporal relations. But although the stimulus generated by this algorithm is clearly perceived as a walking human body, it does not fully capture the true dynamics of human body motion (Saunders, Suchan, & Troje, 2007). Future studies could extend our paradigm to a wider range of naturally generated biological motion. Analogously, a wider range of more natural auditory stimuli could also be studied, and might extend the present focus on footsteps to other cases in which biological motion is associated with auditory events (e.g., hammering a nail, drumming—see Schutz & Lipscomb, 2007).

In conclusion, our experiments used biological-motion stimuli in a multisensory setting and showed clear psychophysical advantages for audiovisual comparisons of temporal frequency when the gestalt of an upright point-light walker was synchronized to auditory events that corresponded naturally with the walker's footsteps. Even in the context of a purely temporal task (for which audition is often considered dominant), the nature of the visual stimuli can substantially affect multisensory judgments, which may additionally be constrained by the ecological validity of the particular audiovisual pairings.

Acknowledgments—A.P. Saygin was supported by European Commission Marie Curie Award FP6-025044 to A.P. Saygin; National Science Foundation (NSF) CAREER Grant 0133996 to V.R. de Sa; NSF Grant BCS 0224321 to M.I. Sereno; and National Institutes of Health Grant RO1 DC0021 to E. Bates. J. Driver was supported by the Medical Research Council (United Kingdom) and by a Royal Society-Leverhulme Trust Senior Research Fellowship. V.R. de Sa was supported by NSF CAREER Grant 0133996. We thank S. Kennett and M.I. Sereno for providing helpful comments; E. Bates for providing testing facilities and support; S. Wilson for helping with experimental stimuli and programs; and M. Datlow, L. Jones, and L. Gordon for assisting with testing.

REFERENCES

- Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences, USA, 102*, 2244–2247.
- Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*, 260–268.
- Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., & Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience, 17*, 377–391.
- Beintema, J.A., Oleksiak, A., & van Wessel, R.J. (2006). The influence of biological motion perception on structure-from-motion interpretations at different speeds. *Journal of Vision, 6*, 712–726.
- Bertenthal, B.I., Proffitt, D.R., & Kramer, S.J. (1987). Perception of biomechanical motions by infants: Implementation of various processing constraints. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 577–585.
- Bertenthal, B.I., Proffitt, D.R., Spetner, N.B., & Thomas, M.A. (1985). The development of infant sensitivity to biomechanical motions. *Child Development, 56*, 531–543.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47–73.
- Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia, 45*, 523–530.
- Calvert, G., Spence, C.J., & Stein, B.E. (2004). *The handbook of multisensory processing*. Cambridge, MA: MIT Press.
- Cutting, J. (1978). A program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods, Instruments, & Computers, 10*, 91–94.
- Grossman, E., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*, 1167–1175.
- Guttman, S., Gilroy, L., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal structure. *Psychological Science, 16*, 228–235.
- Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalarmably poor. *Vision Research, 45*, 1935–1943.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.
- Lewkowicz, D.J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*, 795–804.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.
- Noesselt, T., Rieger, J., Shoenfeld, M., Kanowski, M., Hinrichs, H., Heinze, H.K., & Driver, J. (in press). Audio-visual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*.
- Pavlova, M., Lutzenberger, W., Sokolov, A., & Birbaumer, N. (2004). Dissociable cortical processing of recognizable and non-recognizable biological movement: Analysing gamma MEG activity. *Cerebral Cortex, 14*, 181–188.
- Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics, 62*, 889–899.
- Reid, V.M., Hoehl, S., & Striano, T. (2006). The perception of biological motion by infants: An event-related potential study. *Neuroscience Letters, 395*, 211–214.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.
- Saunders, D.R., Suchan, J., & Troje, N.F. (2007, May). *Point-light walkers with and without local motion features for determining direction*. Poster presented at the annual meeting of the Vision Sciences Society, Sarasota, FL.

- Saygin, A.P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*, 2452–2461.
- Saygin, A.P., Wilson, S.M., Hagler, D.J., Jr., Bates, E., & Sereno, M.I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience*, *24*, 6181–6188.
- Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*, 888–897.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research: Cognitive Brain Research*, *14*, 147–152.
- Shipley, T.F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, *14*, 377–380.
- Spelke, E.S., Smith Born, W., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. *Perception*, *12*, 719–732.
- Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford, England: Oxford University Press.
- Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, *13*, R519–R521.
- Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception*, *13*, 283–286.
- Tadin, D., Lappin, J.S., Blake, R., & Grossman, E.D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience*, *5*, 1010–1015.
- Troje, N.F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector”? *Current Biology*, *16*, 821–824.
- Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1583–1590.
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audiovisual asynchrony. *Brain Research: Cognitive Brain Research*, *22*, 32–35.
- Welch, R.B. (1999). Meaning, attention, and the “unity assumption” in intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam: Elsevier.
- Wichmann, F.A., & Hill, N.J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313.
- Wichmann, F.A., & Hill, N.J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*, 1314–1329.
- Zampini, M., Shore, D.I., & Spence, C. (2005). Audiovisual prior entry. *Neuroscience Letters*, *381*, 217–222.

(RECEIVED 6/24/07; REVISION ACCEPTED 10/15/07)