In the footsteps of biological motion and multisensory perception:

Judgments of audio-visual temporal relations are enhanced for upright walkers

Running head: Biological motion and multisensory perception

Ayse Pinar Saygin [1,2]

Jon Driver [1]

Virginia R. de Sa [2]

[1] UCL Institute of Cognitive Neuroscience, University College London

[2] Department of Cognitive Science, University of California San Diego

Correspondence to:

Dr AP Saygin

UCL Institute of Cognitive Neuroscience

University College London

17 Queen Square, WC1N 3AR, London, United Kingdom

Tel: (44) 207 679 5570

Fax: (44) 207 813 2835

E-mail: a.saygin@fil.ion.ucl.ac.uk

Abstract

Observers judged whether a periodically moving visual display (point-light 'walker') had the same temporal frequency as a series of auditory beeps (that could coincide with apparent 'footsteps' of the walker). Performance in this multisensory judgment was consistently better for upright point-light walker stimuli, than for inverted or scrambled control stimuli, even though temporal information is unchanged in these latter 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) indicating some specificity to the naturally experienced multisensory relation, not just better temporal perception for upright walkers per se. Our experiments indicate that the Gestalt of the 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 additionally constrained by the ecological validity of the particular audio-visual pairings.

Abstract word count: 149

Manuscript word count: 3733 (excluding references and figure captions, including acknowledgments)

Introduction

Natural perception is inherently multisensory, involving the processing and integration of information from multiple modalities. Considering just the auditory and visual modalities alone, there are many examples of stimuli we encounter frequently that are more often bimodal than unimodal (e.g., moving objects creating predictable noises, speech perception).

While the majority of laboratory experiments on perception focus on one sensory modality at a time, multisensory perception has been a research area of interest to psychologists for several decades and 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 & McDonald, 1976) has used relatively impoverished, simple visual and auditory events and as such, little is known about factors affecting the perception of multisensory temporal relations for more complex, meaningful or ecological stimuli.

Here we studied perception of audio-visual temporal relations for a stimulus-class that is relatively natural, and also taps into extensive ecological experience, yet is also highly controllable; namely 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 review). Most studies of biological-motion perception have utilized unisensory visual stimuli only, with just a few exceptions to date (e.g., Arrighi, Alais, & Burr, 2006; Barraclough et al., 2005; Brooks et al., 2007). By contrast, in natural settings perception of biological movements is often multisensory, accompanied by related inputs in other modalities, notably audition (as when footsteps are heard as well as seen). Here, we consider possible implications for perceiving audiovisual temporal relations between seen point-light walkers and auditory events that might correspond to their ‘footsteps’.

A well-established visual finding is that point-light walkers appear less recognizable and possibly less coherent when inverted (e.g. Pavlova & Sokolov, 2000; Shipley, 2003; Sumi, 1984; Tadin, Lappin, Grossman & Blake, 2002). But note that such inversion will not alter the temporal information contained within the local visual motions. Here we examined whether judgments of auditory timing in relation to visual timing might nevertheless be performed more accurately when (for the same auditory sequences) the multisensory judgments concerned upright rather than inverted point-light walkers (Experiment 1). Establishing this would provide initial evidence that the Gestalt of the upright walker may have crossmodal consequences for audio-visual temporal judgments. Having confirmed this, we sought whether a similar disadvantage occurred for scrambled point-light walkers, another control stimuli that contain the same motions of an upright walker but without the Gestalt of a walking figure (Experiment 2). Finally, we investigated whether the improved performance for upright walkers still holds in the less ecological case of greatly phase-offset visual and auditory stimuli, where the sounds could not easily be heard as footsteps (Experiment 3).

General Methods

Participants were adults aged 18-34 with normal or corrected visual acuity and normal hearing by self-report, and 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 on a uniform black background. Each trial also presented periodic sounds (sequences of beeps, see below). The task was to indicate whether the auditory and visual cycles had matching or mismatching temporal frequency.

To produce the visual stimuli, we implemented Cutting's classic algorithm (Cutting, 1978) for generating point-light walkers, in Matlab (Natick, MA). Although this much-studied algorithm is artificially generated (see Discussion), it has the virtue of giving rise to perception of human action while being highly controllable and periodic. The figures were defined by 11 point-lights, some of which were occluded during the motion trajectory for brief durations (e.g., the elbow dot could disappear behind the 'torso'). The point-light figures subtended approximately 5.5 degrees of visual angle in height when viewed from 60 cm. Point-light animations were presented facing right or left, randomly. The figures did not translate on the screen when 'walking', but remained at the centre of the display (Fig. 1).

The auditory stimuli were sequences of 1000 Hz beeps of 100 msec duration each, presented binaurally over headphones, at rates described below.

Participants were told that on each trial, they would see about a dozen periodically moving white dots on a black background and at around the same time, hear periodically presented "beeps", and that the dot and sound streams would each have a constant periodicity. Their task was to judge whether or not the visual and auditory streams had the same temporal or repetition frequency. They indicated by button-press whether there was a temporal “match” or a “mismatch” between the visual and auditory cycle, regardless of the kind of motion (dot displays) they saw on the screen. The experiment began with 12 trials of practice; by the end of this every participant indicated that they understood the task.

Each trial started with a 500 msec fixation point, after which a point-light animation was presented for 2000 msec at 60 frames/sec, 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 auditory and visual temporal frequency. After each response, a visual feedback cue (green dot = correct; red dot = incorrect) appeared for 500 msec. Testing sessions had multiple blocks comprising 30 trials each. Each of several ‘difficulty levels’ (based on the actual temporal-frequency mismatch between visual and auditory stimuli, see below) was tested in two separate blocks, resulting in 60 trials total per level. The order of blocks was pseudo-randomized (a first block at each level was administered before the second blocks) and counterbalanced between participants. Each testing session lasted 70-90 minutes depending on the duration of breaks participants took between blocks.

The average temporal frequency of the visual stimuli was 2 Hz, where 1 full cycle corresponds to a half-gait cycle, or one footstep. Thus it took on average one second for two complete footsteps (one each with the alternating feet) to be completed. The 2 Hz rate per footstep was an average value since the frequency (and thus the walking speed) was jittered randomly (and continuously) between trials by up to ±0.2 Hz. The visual percept of the walker at these frequencies tends to correspond to natural, biologically plausible walking speeds (Beintema, Oleksiak, & van Wezel, 2006). Since each trial was 2 seconds long, the frequency jitter introduced into the visual stimuli meant that some trials would contain slightly more (or less) cycle multiples than others; but on average, there were 4 footsteps per trial (2 steps with each ‘foot’).

On match trials, the auditory stimuli consisted of tones synchronized with every footstep at the point where the ‘foot’ dot makes apparent contact with an apparent ground surface to reverse direction. Frequency match or mismatch between the auditory and visual stimuli was created as follows: First, the walker’s frequency for the upcoming trial was selected (jittered slightly with a mean of 2 Hz as described above). The sound repetition frequency was either the same as the visual frequency (in “match” trials, which occurred on a random half of trials), or was obtained by selecting a frequency that was lower by 0.01-0.4 Hz (except in Experiment 3, when it was lower by 0.02-0.8 Hz). For example, in a mismatch trial with, say, a 2.11 Hz walker at the testing level ‘0.08 frequency offset’, the sound repetition frequency would be 2.11 minus 0.08 = 2.03. The smaller the offset value, the more difficult it is to detect a mismatch. Hence the actual frequency offset in mismatch trials, between the visual and auditory stimuli, constituted the difficulty levels in this experiment. There were 7 blocked difficulty levels in all experiments (0.01, 0.02, 0.05, 0.08, 0.12, 0.2, 0.4) except Experiment 3, which had 6 difficulty levels (0.02, 0.04, 0.08, 0.12, 0.4, 0.8). On mismatch trials, the sounds always had a lower repetition frequency than the walkers. In the case of a mismatch, 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, due to the relatively short trial length.

Each visual ‘footstep’ was defined as the frame in the animations corresponding to when a single foot dot changed trajectory, as if hitting an apparent surface. This occurred at

10% into each one-step sub-cycle. For example, if the walker took exactly 1 second to complete a two-step cycle (i.e. at 2 Hz), the ‘footsteps’ in that trial would occur at 0.1 and 0.6 seconds.

Insert Fig 1

In Experiments 1 and 2, when the temporal frequencies of the visual and auditory stimuli matched (random half of trials within each block), the sounds corresponded to the ‘footsteps’ of the walker-based stimuli. Whether or not the trial featured matching or mismatching stimuli, the first sound always coincided with the first visual footstep. In Experiment 3, we deliberately phase-offset the sounds away from the footstep frames (by half of the gait cycle), even when matching in frequency.

Stimuli were presented and responses collected using the Psychophysics Toolbox for Matlab (Brainard, 1997). We used bootstrapping methods for estimating each individual participant’s perceptual thresholds, and for comparing 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 differences between thresholds was calculated using similar methods, with the same toolbox plus the *pfcmp* toolbox (Wichmann & Hill, 2001b). We also used standard repeated measures ANOVA (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. Only accuracy data are reported here since sensitivity analyses paralleled accuracy analyses and no further insights emerged.

Experiment 1

Here we contrasted upright and inverted point-light walkers in the paradigm described above. Half the trials contained upright walkers, the other half contained inverted walkers, randomly. Eight adults (5 females) participated and performed a 2AFC task (same or different audio-visual repetition frequencies) on each trial.

The critical result was that participants performed the audiovisual task better with upright walkers than inverted point-light walkers, even though these two situations contain identical information in terms of temporal frequency. Figure 2a shows the data combined from all participants as an overall psychometric function. Better performance on the upright walker condition can readily be seen and was statistically significant. We tested the difference between these two psychophysical curves using bootstrapping (see General Methods), and found that the likelihood that they come from the same underlying distribution was very small ($p = 0.001$). Furthermore, for each of the eight participants individually, the estimated 75% accuracy threshold for upright walkers was significantly less than that for inverted walkers ($p<0.05$, two-tailed), as established via bootstrapping analyses on each individual participant's data.

Insert Fig 2

Experiment 2

Our next experiment sought to generalize the finding from Experiment 1 to the case of upright walkers versus scrambled walkers. Scrambled stimuli have been used extensively as control stimuli for biological motion across a range of methodologies (e.g. Grossman & Blake, 2002; Ikeda, Blake, & Watanabe, 2005; Pavlova, Lutzenberger, Sokolov, & Birbaumer, 2004; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Saygin, 2007).

Half the trials had an upright walker as the visual stimulus, the remaining half a scrambled walker, in a randomly intermingled sequence. Our new control visual stimuli were created by spatial scrambling, i.e., randomizing the starting positions of the walker dots while keeping their local motion trajectories intact. The scrambled walker thus continued to have the same (local) information about temporal frequency as the intact walker – but the relation between visual dots (and thereby the Gestalt) was affected. A single scrambled animation and its mirror image (about the vertical axis) were used throughout the experiment, as were the right or left-facing upright walkers (also mirror-images about the vertical). As before, the first sound in the sequence always coincided with the change in trajectory for one or other 'foot' dot, although in the case of the scrambled walker, these foot dots now appeared in scrambled locations with respect to each other and to the other dots.

One participant from Experiment 1 and three new participants were tested in Experiment 2 (2 females). They performed the audio-visual task significantly better with the upright walker in comparison to the scrambled walker. The overall psychophysical curves for the upright and scrambled stimuli are shown in Figure 2b and differed significantly from each other ($p<0.001$). For three of the four participants, the estimated 75% accuracy threshold for upright walkers was, individually, significantly smaller than their threshold for scrambled walkers ($p<0.05$, two-tailed); in the fourth participant there was a trend in the same direction but it did not reach significance.

Experiment 3

Our final experiment tested whether the inversion effect 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 aligning the first sound on each trial to a visual footstep, for all trials the first sound now occurred 50% into each footstep (i.e. at 25% and 75% in terms of the complete two-step cycle), which never coincided with a visual footstep nor with another dot changing trajectory. The task and procedure were otherwise as in Experiment 1.

If the advantage for intact upright walkers reflects some nonspecific general advantage in their temporal perception within vision alone, then we might again expect improved performance for upright walkers as compared with inverted walkers. If instead the advantage observed for intact upright walkers reflects the (natural) temporal correspondence between the Gestalt of the visual walker and the auditory footsteps, then the advantage for upright intact walkers should now diminish or disappear.

One participant from Experiment 1 and three new participants were tested (2 females). The results were qualitatively different, now 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 with a one tailed test, p=0.4, indicating no difference). There were no significant differences between the two conditions in any of the individual participants' estimated thresholds either (even with one-tailed tests, all p's>0.1). To confirm the difference in outcome for Experiments 1 and 3 more formally, we conducted a statistical comparison between these experiments of the inverted-minus-upright differences in thresholds (n=8 for Experiment 1, n=4 for Experiment 3). The inversion effect was indeed significantly larger in Experiment 1 than the (null) effect in Experiment 3, as confirmed both with a between-subject F-test (F(1,11) = 12.25; p = 0.003, one-tailed) and also by a nonparametric Mann-Whitney U-test (z = -2.717; p = 0.003, one-tailed).

This final experiment confirms that the advantage found for intact upright walkers is specific to the situation of natural correspondence between seen walking and heard 'footsteps', rather than being due to a nonspecific advantage for upright visual 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, or scrambling in Experiment 2), the audio-visual comparison was impaired. The advantage for upright versus 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 phase-locked does not simply reflect better perception for upright visual walkers per se, but rather a benefit specifically when the sounds are synched with particular visual events that could plausibly produce them (as for heard footsteps).

A previous purely visual study (Tadin, Lappin, Blake, & Grossman, 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 crossmodal effect on the particular phase-relation of auditory and visual footsteps (Experiments 1 vs. 3) indicates that the benefit for intact upright walkers here may not be purely visual, but 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, as for heard footsteps.

Recent experiments by Guttman et al (2005) with very different stimuli showed that visual 'rhythms' may be automatically encoded in the auditory domain, which may arguably specialize in processing temporal structure (Welch, 1999). But note that here the visual Gestalt of the walker played a critical role, not just the temporal structure per se (which was the same for the inverted or scrambled point-light walkers that led to impaired judgments). A further factor to consider is that perception of body movements can engage the viewer's own motor system, even for point-light stimuli (see Saygin et al., 2004; Saygin, 2007 for neuroimaging and neuropsychological evidence); thus the temporal structure might also become encoded motorically when biological motion is perceived. Future variations on our audio-visual paradigm, measuring neural activity via imaging, may shed light on the brain systems affected, and whether these include regions conventionally linked to biological motion (e.g. STS, Grossman & Blake, 2002; Barraclough et al., 2005; Saygin et al., 2004), modality-specific auditory and visual areas (e.g. Noesselt et al., in press) and/or motor components of the so-called 'mirror' system (Rizzolatti & Craighero, 2004).

We have suggested that our results reflect a crossmodal advantage when the Gestalt of an intact upright walker is seen with a natural temporal relation to heard footfalls. Recently Troje & Westhoff (2006) have shown that foot-dots convey special information in biological-motion displays and that they may be particularly important for inversion effects. One might wonder for our paradigm, whether the entire walker is indeed necessary, or just the foot-dots. In a control study with four new participants (not reported in full here), we found that the advantage of upright over inverted presentation disappeared when only the two foot-dots were shown. Further research with variants on our paradigm could establish exactly which aspects of the upright walker are necessary to produce our crossmodal effect.

Our results bring together traditionally separate psychological domains (perception of biological motion, and mutisensory integration) while suggesting further research directions for each. We found enhanced performance when the visual walker and its footsteps corresponded temporally (in both frequency and phase) to the auditory footsteps. Such an association between seen and heard footsteps will often be observed in daily life, although presumably with slight variations in audio-visual offset (e.g. slightly different auditory delays depending on viewing distance – see Arrighi, Alais, & Burr, 2006; Spence & Squire, 2003). Future studies might examine whether perfect synchrony for the matching condition is optimal, or whether some slight fixed auditory delay can be dealt with or even adapted for (Vroomen, Keetels, de Gelder, & Bertelson, 2004).

The present 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 the perception of audio-visual temporal correspondences in infants (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 literature bringing these two topics together, as we sought to accomplish here in adults. Developmental variants on the present paradigm (e.g. selective looking or crossmodal habituation in infants) might provide a useful way to test whether the present crossmodal effect requires hard-wired mechanisms for detecting audio-visual correspondence, or instead experience with particular types of real-world stimuli in which such correspondence arises.

As mentioned in the introduction, our study sought to go beyond prior work on crossmodal 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. The Cutting (1978) algorithm used for our visual stimuli is well established in the field, and had the virtue of allowing precise control of temporal relations. But while it is clearly perceived as a walking human body, it does not fully capture the true dynamics of human body motion (Saunders, Suchan, & Troje, 2007, May). 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 where biological motion is associated with auditory events (as when hammering a nail, drumming, etc – see Schutz & Lipscomb, 2007).

In conclusion, our experiments used biological motion stimuli in a multisensory setting, and showed clear psychophysical advantages for audio-visual comparisons of temporal frequency in cases where the Gestalt of an upright point-light walker was synched 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 audio-visual pairings.

Acknowledgments: APS was supported by European Commission Marie Curie award FP6-025044; NSF CAREER grant 0133996 to VRD; NSF BCS 0224321 to M.I. Sereno; and NIDCD RO1 DC0021 to E. Bates. JD was supported by the Medical Research Council (UK) and a Royal Society-Leverhulme Trust Senior Research Fellowship. VRD was supported by NSF CAREER grant 0133996 (USA). We thank S. Kennett and M.I. Sereno for helpful comments; E. Bates for providing testing facilities and support; S. Wilson for help with experimental stimuli and programs; and M. Datlow, L. Jones and L. Gordon for assistance 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 Wezel, 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.

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.

Ikeda, H., Blake, R., & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision Research, 45*, 1935-1943.

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 and 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 Vision Sciences Society meeting, Sarasota, FL.

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.

Saygin, A.P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130*, 2452-2461.

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.

Spence, C.J, & Driver, J. (2004). *Crossmodal space and crossmodal attention.* Oxford: Oxford University Press.

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., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*, R519-521.

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 audio-visual 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 and Psychophysics, 63*, 1293-1313.

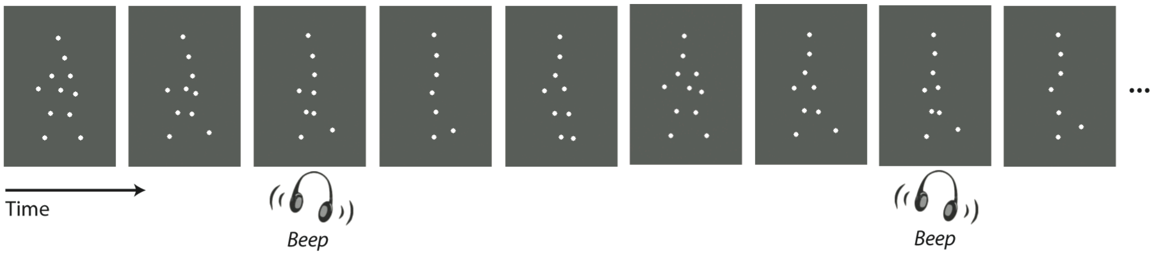
Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics, 63*, 1314-1329.

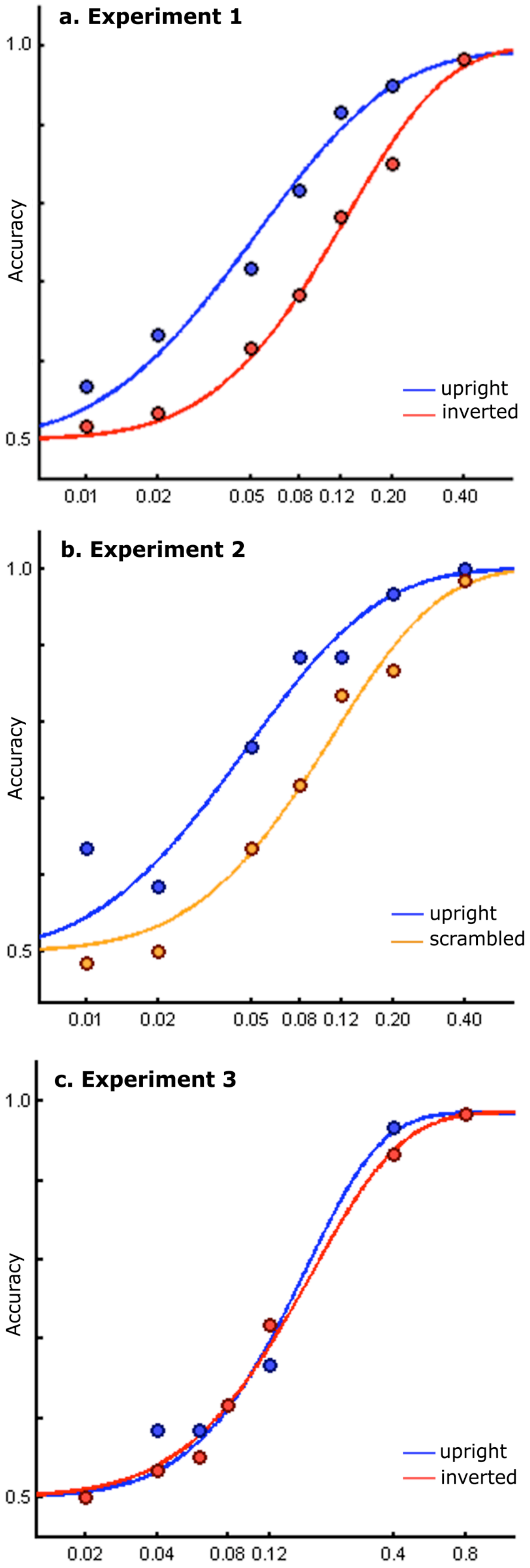
Zampini, M., Shore, D. I., & Spence, C. (2005). Audiovisual prior entry. *Neuroscience Letters, 381*, 217-222.

Figure captions

Figure 1. Schematics of the stimuli. Example frames from upright point-light walkers are depicted for the case of audiovisual match, in which the sounds were played simultaneously with frames in the animations where one or other of the feet dots appeared to hit the ‘ground’ and change movement trajectory (the ‘footstep frames’).

Figure 2. Psychophysical curves depicting data from all subjects in (a) Experiment 1, (b) Experiment 2 and (c) Experiment 3. The x-axis denotes the frequency difference between the sound and the visual motion in mismatch trials (in Hz, plotted on a log scale) and corresponds to task difficulty (see General Methods). The y-axis denotes accuracy. Individual psychophysical thresholds were also calculated and analyzed (reported in the text).


Time
Beep
Beep
...


a. Experiment 1
1.0
0.5
Accuracy
0.01
0.02
0.05
0.08
0.12
0.20
0.40
upright
inverted
b. Experiment 2
1.0
0.5
Accuracy
0.01
0.02
0.05
0.08
0.12
0.20
0.40
upright
scrambled
c. Experiment 3
1.0
0.5
Accuracy
0.02
0.04
0.08
0.12
0.4
0.8
upright
inverted