

Perceptual and computational analysis of critical features for biological motion

Steven M. Thurman

Department of Cognitive Science, Center for Cognitive Neuroscience, University of California, Irvine, USA



Section for Computational Sensomotrics, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research and Centre for Integrative Neuroscience, University Clinic, Tübingen, Germany

Martin A. Giese



Emily D. Grossman

Department of Cognitive Science, Center for Cognitive Neuroscience, University of California, Irvine, USA



Among the most common events in our daily lives is seeing people in action. Scientists have accumulated evidence suggesting humans may have developed specialized mechanisms for recognizing these visual events. In the current experiments, we apply the “bubbles” technique to construct space–time classification movies that reveal the key features human observers use to discriminate biological motion stimuli (point-light and stick figure walkers). We find that observers rely on similar features for both types of stimuli, namely, form information in the upper body and dynamic information in the relative motion of the limbs. To measure the contributions of motion and form analyses in this task, we computed classification movies from the responses of a biologically plausible model that can discriminate biological motion patterns (M. A. Giese & T. Poggio, 2003). The model classification movies reveal similar key features to observers, with the model’s motion and form pathways each capturing unique aspects of human performance. In a second experiment, we computed classification movies derived from trials of varying exposure times (67–267 ms) and demonstrate the transition to form-based strategies as motion information becomes less available. Overall, these results highlight the relative contributions of motion and form computations to biological motion perception.

Keywords: biological motion, computational modeling, bubbles, reverse correlation, classification images

Citation: Thurman, S. M., Giese, M. A., & Grossman, E. D. (2010). Perceptual and computational analysis of critical features for biological motion. *Journal of Vision*, 10(12):15, 1–14, <http://www.journalofvision.org/content/10/12/15>, doi:10.1167/10.12.15.

Introduction

Everyday we see people moving about the world and interacting with other individuals. These events form the basis of our daily social interactions, and perhaps as a result, there is evidence to suggest that humans have developed specialized cortical mechanisms for perceiving biological movement patterns. This view is supported, in part, by the observation that human observers spontaneously recognize human activity from sparse point-light (PL) depictions of actions that consist of only a handful of markers attached to the head and joints of the body (Johansson, 1973). Since the adoption of these stimuli by vision science, researchers have demonstrated the apparent ease with which observers recognize these sequences as human actions.

Since this early work, there have been a number of technological advances that have served to cultivate a large body of research on biological motion perception. It has been shown, for instance, that human observers can recognize a large number of actions in addition to

locomotion, including instrumental actions and social interactions (Dittrich, 1993). Observers accurately discriminate identity (Cutting & Kozlowski, 1977; Troje, Westhoff, & Lavrov, 2005), gender (Kozlowski & Cutting, 1977; Troje, 2002), and emotion (Roether, Omlor, Christensen, & Giese, 2009) from PL displays of dance (Dittrich, Troscianko, Lea, & Morgan, 1996) and arm movements (Pollick, Paterson, Bruderlin, & Sanford, 2001). Clearly, the sparse information in PL animations is sufficient for nuanced discriminations of movement styles.

Perception is also robust to many manipulations that degrade the information available in PL animations. Observers can recognize human activity when very short PL animations are embedded in motion-matched clutter, or noise masks (Thirkettle, Benton, & Scott-Samuel, 2009; Thurman & Grossman, 2008). Blurring the dots, randomizing the contrast polarity, and presenting subsets of the dots dichoptically also does not impair performance (Ahlström, Blake, & Ahlström, 1997, but see Aen-Stockdale, Thompson, Hess, & Troje, 2008). Further, recognition does not depend on the exact placement and continuous visibility of the tokens, as demonstrated by

accurate discrimination of PL animations created with the positions of dots jittered randomly along the limb and with an imposed limited lifetime (Beintema & Lappe, 2002).

However, there have been many reports on the limitations of PL biological motion perception. One of the earliest observations by Johansson (1973) was that naive observers do not report stationary frames of PL sequences to be biological, even though they are recognized as biological when set into motion. This would seem to imply a critical importance of motion analysis in biological motion recognition. PL perception is also impaired under dim light conditions and completely disrupted when the PL markers and background are displayed at isoluminance, two manipulations that impair visual analyses of complex motion patterns (Garcia & Grossman, 2008; Grossman & Blake, 1999). Scrambling dot positions or jittering the spatiotemporal phase relations of PL dots seriously disrupts perception (Ahlström et al., 1997; Bertenthal & Pinto, 1994; Troje & Westhoff, 2006). Together, these findings highlight the importance of spatiotemporal cues in the perceptual construction of PL biological motion.

From the current evidence in the literature, researchers have developed competing computational models (often built from statistical and psychophysical measurements of dominant features in the PL stimuli) that differ in the extent to which PL biological motion recognition depends on motion and form analyses. For example, Troje (2002) used principal components analysis to decompose the spatiotemporal features that differed between male and female PL walkers. From these features, new PL sequences were constructed with gender defined only by select subsets of the features. It was found that observers relied more strongly on motion cues than form cues, performance that was replicated with a linear classifier. The conclusion was that much of the nuanced gender information carried in the PL displays could be characterized by a limited number of motion features.

The importance of motion features has also been highlighted by other computational studies using biological motion. Motivated by a physiologically inspired computational model that reproduces a variety of experimental data from the literature (Giese & Poggio, 2003), Casile and Giese (2005) analyzed the visual features available in two separate computational pathways of the model corresponding to form and motion analyses. The results from this model revealed a stronger influence of motion features for the perception of PL stimuli. Furthermore, a principal components analysis on mid-level form and motion cues in PL walkers showed that only the motion features were largely shared between stick figure (SF) and PL representations, more evidence for the importance of motion analysis. Artificial PL animations built with these critical mid-level motion features, the sinusoidal antiphase horizontal motion of the wrists and ankles, were also accurately perceived as human walkers (Casile & Giese, 2005). Separate experiments also

suggested the importance of this motion feature, finding that visual sensitivity to point-light actions fluctuated sinusoidally across time, with peak performance during the moments of antiphase horizontal, or opponent, motion (Thurman & Grossman, 2008).

There has been other psychophysical evidence, however, that has emphasized the importance of form cues for biological motion perception. For example, expert observers can learn to discriminate point-light movements from a single body posture, which can be taken as evidence for form analysis (Thirkettle et al., 2009; Todd, 1983). Likewise, biological motion suffers from an inversion effect, whereby displaying PL animations upside-down impairs accurate recognition of actions (Dittrich, 1993; Troje & Westhoff, 2006), emotions (Dittrich et al., 1996), gender (Barclay, Cutting, & Kozlowski, 1978), and detection in noise masks (Bertenthal & Pinto, 1994; Pavlova & Sokolov, 2000). In face perception, the inversion effect has been argued to be evidence for holistic processing (Farah, Wilson, Drain, & Tanaka, 1998), and the inversion effect for bodies has been taken as evidence for the use of global form templates as a means for recognition (e.g., Reed, Stone, Bozova, & Tanaka, 2003). An emphasis on form analysis is also integral to a number of computational models that discriminate biological motion on the basis of form templates (e.g., Marr & Vaina, 1982; O'Rourke & Badler, 1980).

In more recent computational models, a template-matching procedure compares input frames of a PL animation sequence to stored templates of stick figures. The template-matching approach is based exclusively on form information and also replicates a number of psychophysics and neuroimaging results from the literature (Lange, Georg, & Lappe, 2006; Lange & Lappe, 2006).

The current experiments seek to clarify the discrepancies in the literature as to the key features and computational mechanisms of biological motion perception. Our previous experiments using the temporal “bubbles” method found key temporal intervals in PL walking direction discriminations, and when discriminating biological from non-biological motion (Thurman & Grossman, 2008). These current experiments used spatiotemporal “bubbles” (Fiset et al., 2008; Vinette, Gosselin, & Schyns, 2004) to characterize the spatial and temporal nature of the diagnostic features that human observers use when discriminating biological motion. We measured this for two depictions of human actions, point-light (PL) and stick figure (SF) walkers. This method used reverse correlation with Gaussian aperture masks to compute classification movies that illustrate the significant spatial and temporal regions of the stimulus used for discriminating the human walkers. If classification movies between PL and SF depictions were found to be similar, it would imply that critical features are invariant across the two depictions of the biological motions. This invariance might reflect mid-level motion features that are largely shared between the two depictions (Casile & Giese, 2005). Alternatively,

differences between the classification movies might indicate how computational and perceptual strategies differ when form is made explicit in SF as compared to when it is implicit in PL depictions.

Next, to clarify the role of motion and form computations for biological motion perception, we computed classification movies from the performance of a previously defined model of perception (Giese & Poggio, 2003) using as input the same “bublized” walker stimuli that were shown to observers. The architecture of this model mimics fundamental features of the human visual system and contains separable form and motion pathways that analyze basic shape features (i.e., orientation) and optic flow, respectively. The pathways are organized hierarchically, with more complex representations and greater size and position invariance at higher levels. We compared the model-derived classification movies for SFs and PL figures and analyzed the spatiotemporal diagnostic features that modulate model responses in each pathway. These results were quantitatively compared to behavioral classification movies, drawing a link between human behavior and computations in this neural model of perception.

Finally, we report on a second experiment that analyzed computational strategies as observers were exposed to shorter durations of biological motion (down to 66 ms), which limited the quality and amount of motion information while leaving stationary cues intact, albeit for shorter periods of time. We hypothesized that observers would shift strategies as a function of exposure time, and such shifts would be manifested in the classification movies. We also tested the model’s performance as a function of exposure time and again compared the behavioral and model classification movies.

Experiment 1

Participants

Twenty (15 females, 5 males) unpaid undergraduate students were recruited from the University of California, Irvine research participation pool and received course credit for participation. None of the participants had prior experience with PL biological motion and were naive to the purpose of the experiment. Participants reported normal or corrected-to-normal vision and gave written consent in accordance with the University’s IRB protocol.

Materials

Participants were seated 40 cm from one of five CRT monitors (refresh 60 Hz) in the Social Sciences Research Laboratory at UC Irvine controlled by Pentium 4 Dell

computers running Windows XP. The experiment was programmed in Matlab (version R2008a) using functions from the Psychophysics Toolbox (version 3.0.8; Brainard, 1997; Pelli, 1997). The stimulus display rate was 30 frames per second (33.3 ms/frame).

The PL walker stimulus was recorded from an actor using a VICON 512 motion capture system and is the same as used in previous experiments (Jastorff, Kourtzi, & Giese, 2006; Thurman & Grossman, 2008). It consisted of 13 markers representing the head plus the left and right shoulders, hips, elbows, wrists, knees, and ankles. The PL markers were represented by small dark squares (0.2 deg) on a light gray background. Horizontal translational components were removed so that the walker appeared to be walking on a treadmill in the profile, or sagittal view. Leftward and rightward walkers were created from the same animation sequence, mirrored across the vertical axis. The SF was created from the PL actor by drawing solid lines (approximately 0.1 degree of visual angle in thickness) between the locations of the markers, consistent with the human skeleton. Each stimulus consisted of a complete 2.1-s gait cycle, completed over 62 individual frames, or postures, and subtended approximately 7.6 degrees visual angle vertically and 4.8 degrees horizontally.

Masked, or “bublized”, biological motion is shown in Figure 1A. This stimulus was created by first randomly selecting an interval of the gait cycle (467 ms or 14 frames), and then applying a bubble mask to each frame in the sequence. The bubble mask was mostly opaque but revealed portions of the stimulus through randomly distributed Gaussian apertures with a constant diameter of 0.8 deg. Thus, the PL or SF animations were largely invisible, except for the subsets that were visible within the small windows of the mask. The locations of the Gaussian apertures were chosen randomly and independently from trial to trial and were stationary throughout the trial sequence.

On each trial, the bubblized interval of biological motion was inserted into a larger centrally located stimulus aperture (12 × 12 deg) and was jittered randomly up to 4 deg in each direction from center. The onset of the target stimulus in the trial sequence varied between 33 and 167 ms after the appearance of the gray stimulus aperture, which signified the start of the trial.

Procedure

Subjects completed 1500 PL or SF bubble trials in a 1.5-h behavioral session. The number of bubbles on each trial was adjusted online using a 3–1 double interleaved staircase procedure (Levitt, 1970) to maintain 80% threshold accuracy in discriminating the walking direction of the figure (left versus right). Every incorrect response led to the addition of one bubble to the mask, making more spatiotemporal information visible to the observer and thus the task easier. Three consecutive correct responses

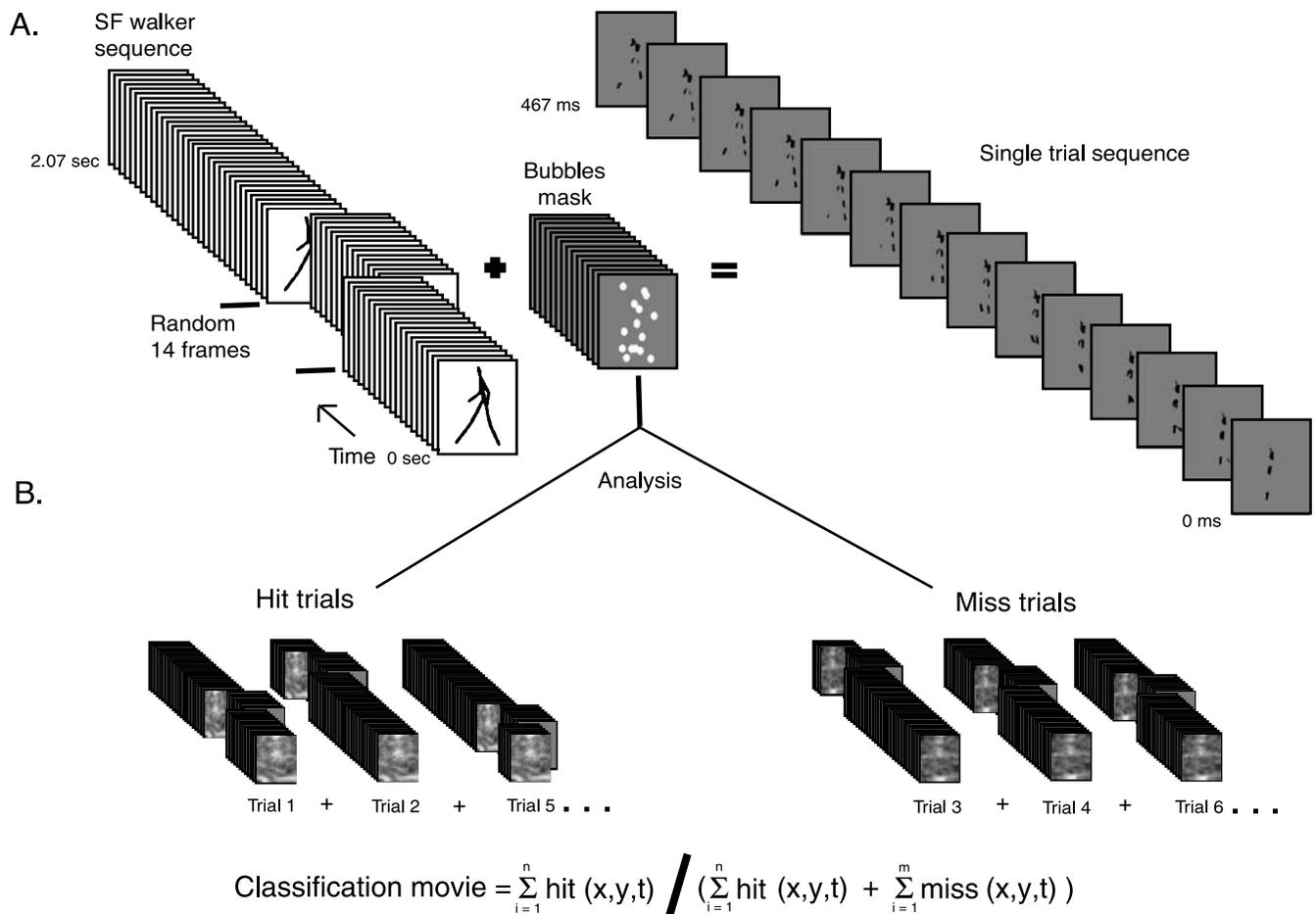


Figure 1. Schematic of the “bubbles” stimulus construction and method of analysis. (A) A sample trial from [Experiment 1](#) includes randomly selecting 14 frames of the walker (SF shown) and masking with stationary Gaussian windows (bubbles) to yield a 467-ms sequence in which partial stimulus information is visible. (B) Analysis: If the trial was judged correctly in the left–right discrimination task, then the spatiotemporal bubble mask is added to the volume of “hit” trials. If judged incorrectly, then the mask is added to the volume of “miss” trials. Classification movies are computed by dividing the hit volume by the volume of hits + misses, and then normalizing as described in the text.

led to one bubble being removed from the image mask, making the task more difficult. Adjusting the number of bubbles to achieve a target performance criterion is the standard procedure introduced by Gosselin and Schyns (2001), and allowing the number of bubbles to fluctuate helps make it a “self-calibrating” technique. The critical factor modulating performance from trial to trial is the specific spatiotemporal features revealed by the locations of the randomly distributed bubbles in the mask, and not necessarily the number of bubbles in the mask. In other words, the threshold number of bubbles is essentially an estimate of the percentage of the stimulus space that must be sampled, on average, to achieve threshold performance.

Each participant completed three blocks of 500 trials and indicated their responses by pressing the appropriate key on a keyboard. The staircase was initiated with 15 bubbles in the first block, and subsequent blocks started with the staircase estimate of the threshold number of bubbles

from the previous block. The staircase procedure reached threshold fairly quickly for each participant (after about 70 trials, on average), and the number of bubbles was relatively stable once threshold was reached.

Each subject was randomly assigned to perform the task with either all SFs or PL figures (10 subjects for each condition). Prior to participation in the main experiment, subjects viewed four cycles of an unmasked walker, and all immediately identified it verbally as biological (i.e., a person walking). Participants were then shown five samples of the “bubbled” target stimulus during task instructions.

Analysis

The threshold number of bubbles was estimated for each observer by averaging the number of bubbles across all trials in the final block of the experiment (500 trials).

Using reverse correlation, we computed classification movies from the bubble masks, with trials sorted by observer accuracy. The masks were first put into common body-centered coordinates (the leftward walker) by reflecting the masks from rightward walking trials across the vertical meridian. We then performed a multiple linear regression on the bubble masks across trials (explanatory variables), using the observer response accuracy as predictor variables. For computational simplicity, this amounted to summing all space–time bubble masks that led to a correct response and dividing by the sum of all masks across all trials, including hits and misses. The result was a volume of regression coefficients that represented, for each pixel and time point, the probability of a hit when revealed to the observer. We performed this analysis on all the subject data combined (15,000 trials) for each condition, resulting in a single spatiotemporal classification movie for SFs and PL figures.

To assess statistical significance, we transformed each classification movie into z -scores by estimating the mean and standard deviation of the null distribution from background regions of the stimulus sequence that never contained a signal pixel. This was done for each frame of the stimulus separately, and the resulting maps were then spatially smoothed with a Gaussian filter ($SD = 4$ pixels). Finally for visualization (Figure 2), we averaged across identical postures in the gait cycle (1 gait cycle = 2 steps), threshold the maps at $Z_{crit} = 1.65$, $p < 0.05$ (uncorrected), and overlaid the walker stimulus. To correct for multiple comparisons, we also conducted a pixel test ($Z_{crit} = 3.84$, $p < 0.05$, indicated in Figure 2 as black lines on color bars) using the Stat4Ci toolbox (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005), which is optimized for use with classification images.

Finally, to quantify the similarity between classification movies, we computed the linear correlation coefficient between the derived classification movies. This is a widely used approach for computing similarities between two images for image registration, object recognition, as well as with salience maps generated from human eye movement data (Le Meur, Le Callet, Barba, & Thoreau, 2006). We constrained this similarity analysis to regions of the image with potential stimulus information on each frame, with the intent of minimizing the effect of noisy data points from the peripheral background. This was implemented as including those pixels with PL dots, as well as a 20-pixel diameter circle around each dot, for a total of 185,964 pixels included across all 62 frames. The circle size was chosen to match the size of a single bubble in the experiment (20-pixel diameter). Similarly, for the SF condition we drew a 20-pixel diameter silhouette around the stick figure and included only those pixels in the analysis, for a total of 256,860 pixels. To estimate statistical significance from the large number of paired samples in this analysis, we performed a randomization test that randomly permuted the mappings between all

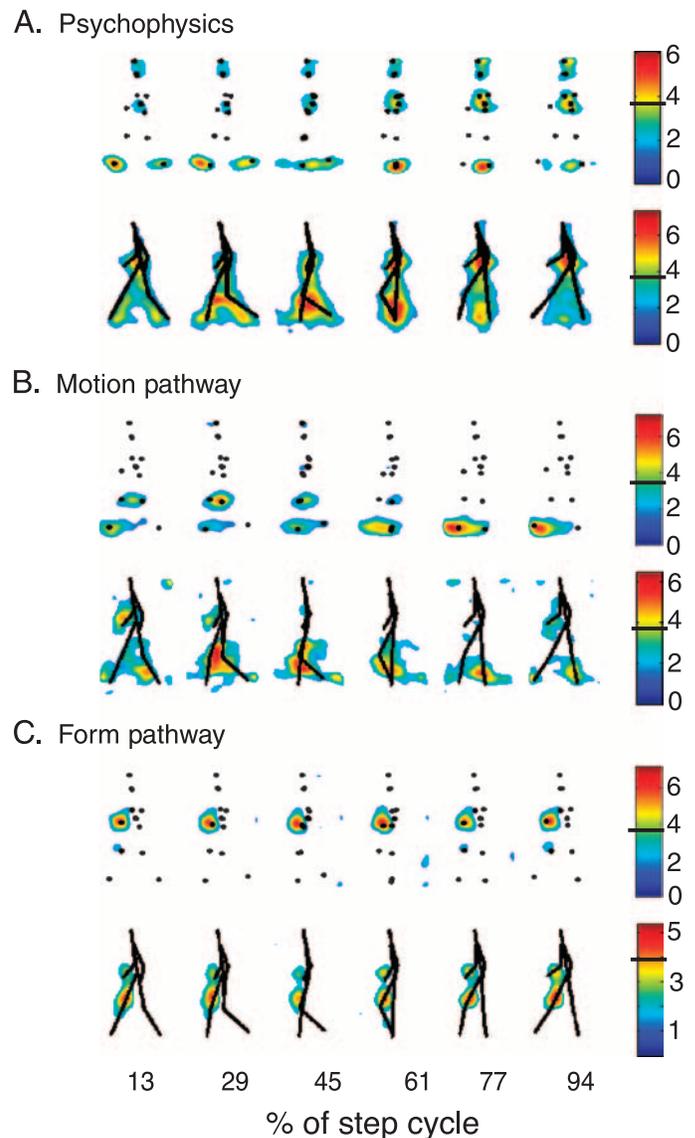


Figure 2. Classification movie results from Experiment 1. (A) Classification movies derived from the human observers for the (top) PL and (bottom) SF conditions. (B) Classification movies derived from the model performance, restricted to the motion pathway. (C) Classification movies derived from the model's form pathway. For visualization, we apply a threshold ($Z_{crit} = 1.65$, $p < 0.05$, uncorrected), such that only significant pixels (i.e., those that deviate significantly from chance) are colored according to the respective color bars. The black lines on the color bar indicate the critical Z -score for the corrected pixel test ($Z_{crit} = 3.84$, $p < 0.05$ corrected). The data are visualized on selected frames of a single step cycle.

data points in the classification movies in 1,000 independent simulations. The computed correlation coefficients from the simulations resulted in an estimate of a null distribution for each stimulus type (PL and SF). The null distribution for PL and SF movies had a mean correlation

	Threshold		Correlation coefficient	
	PL	SF	PL	SF
Psychophysics	33.3	9.1	0.312	0.312
Motion pathway	28.0	42.6	0.100	0.079
Form pathway	32.8	23.4	−0.125	0.017

Table 1. Summary of bubble thresholds and correlation results for all conditions in [Experiment 1](#). The correlation coefficients for psychophysics represent the correlation between PL and SF classification movies. The correlation values for each pathway represent the correlation between the model classification movies for each stimulus condition and the respective psychophysics classification movies.

coefficient of 0, with a standard deviation of 0.0024 and 0.0020, respectively.

Results

Observers required on average 33.3 ($SD = 9.3$) bubbles for discriminating the PL figures, and 9.1 ($SD = 1.7$) bubbles for discriminating the SF figures. Bubble thresholds are also reported in [Table 1](#).

The results of the classification movie analyses are shown in [Figure 2A](#), with animated sequences in [Supplementary Movies 1](#) and [2](#). Subjects used different spatial patterns of diagnostic cues at different times within the gait cycle, but similar spatiotemporal features across SF and PL representations. The correlation coefficient between the classification movies was highly significant, $r = 0.312$, $p < 0.001$. The most significant features for both conditions were contained within the feet and the arms, depending on the phase of the gait cycle. At the moment when the gait is in full stride, the upper body was the most diagnostic for the left–right discrimination task. When the ankles cross the midline, the feet were the most diagnostic. These cues are likely the most potent because they contain the strongest acceleration cues and the most relative (opponent) movement of the limbs, respectively (Casile & Giese, 2005; Chang & Troje, 2009). Due to the lack of dynamic cues in the head and shoulders, these upper body regions likely reflect the use of structural form-based cues.

These results illustrate two important points. First, subjects tapped into different spatiotemporal features depending on the phase of the gait cycle. In some instances, the most diagnostic features were contained in the lower body, and in other instances, they were in the upper body. Thus, subjects can be flexible in their discrimination strategy depending on the limitations of the features available. Second, the overall spatiotemporal pattern of diagnostic information for SF and PL representations was extremely similar, suggesting that these

cues are reliable when form features are depicted implicitly in PL figures or explicitly in stick figures.

Model

To test the role of motion and form computations, we computed space–time classification movies from the performance of a neural model of biological motion perception (Giese & Poggio, 2003). This feedforward model was chosen because it has a biologically plausible architecture that reproduces several experimental results on biological motion perception, and it contains separate computational pathways that mimic the functional divisions of form and motion processing in the ventral and dorsal visual pathways, respectively (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982).

By computing the performance of these two pathways independently, we tried to obtain a measure of the information that is relevant in these two cues to compare to the behavioral data. Below is an abbreviated summary of the model. Additional details, as well as a schematic sketch of the model, are available in previously published reports (Casile & Giese, 2005; Giese & Poggio, 2003).

The purpose of the computational form pathway is to extract the posture-based features of biological motion, even when specified as PL animations. This component of the model initially detects local orientation features of the image, much like simple cells in primary visual cortex (Hubel & Wiesel, 1962). The second level of the form hierarchy pools across these orientation features, much like complex cells in V1 (Hubel & Wiesel, 1962) and neurons in V2 and V4 (Gallant, Connor, Rakshit, Lewis, & van Essen, 1996; Hegde & van Essen, 2000). The result of this pooling operation is orientation detectors with partial scale and position invariance, whose output signals serve as input to detectors of complex shapes in the next level of analysis. These shape detectors are arranged to detect patterns that look like body configurations, or “snapshots”, and they model view-tuned complex shape-selective neurons that have been found in the inferotemporal (IT) cortex of monkeys (Logothetis & Sheinberg, 1996; Tanaka, 1996). These neurons were modeled as radial basis functions, whose centers were defined through training with sample shapes that were derived from movement sequences. These “snapshot neurons” respond selectively to key body postures arising during the biological motion stimuli. The highest level of the hierarchy is comprised of complex motion pattern detectors that temporally smooth and summate the activity of the snapshot neurons that were trained on a particular movement sequence. Such neurons might be present, for example, in the superior temporal sulcus of monkeys (Puce & Perrett, 2003; Vangeneugden, Pollick, & Vogels, 2009).

The motion pathway computes motion energy in the action sequences based on an analysis of the local motion energy, or optic flow, in the animations. In the brain, this computation is likely realized by motion-selective neurons in area MT (Smith & Snowden, 1994). For computational efficiency in this instantiation of the model, optic flow was computed from pixel map sequences using the Horn–Schunck algorithm (Horn & Schunck, 1981). We also tested the Lucas–Kanade method (Lucas & Kanade, 1981) and obtained very similar optic flow sequences, suggesting the choice of algorithm in this case does not substantially influence model performance. To improve the optic flow estimates and reduce the negative influence of the aperture problem, we computed optic flow for each limb separately and then combined optic flow estimates across the limbs using the max operator for each pixel and time point. This special computation is similar to the computation of a layered optic flow (Wang & Adelson, 1994) and ensures correct motion vectors specifically at occlusion points. This step was motivated by the observation that humans seem able to generate correct local motion estimates for articulated figures, as opposed to standard optic flow algorithms that typically produce erroneous results at occlusion points. The next level of the motion hierarchy is designed to be sensitive to translational motion and motion edges, or opponent motion, much like neurons found in area MT and subregions of the MST (Allman, Miezin, & McGuinness, 1985; Tanaka, Fukuda, & Saito, 1989). The third level uses these mid-level motion signals as input for radial basis functions that recognize characteristic instantaneous optic flow patterns that arise during the body motion stimuli. The neural detectors are the equivalent of the snapshot neurons in the form pathway. The final level consists again of motion pattern neurons that temporally smooth and summate activity from the radial basis function units that encode optic flow patterns belonging to the same type of movement sequence. Finally, in both pathways temporal sequence selectivity is implemented with recurrent neural networks that have asymmetric lateral connections (Giese & Poggio, 2003).

In implementing this model, we first trained a single pathway (motion or form) with full sequences of SF and PL figures walking left and right. Most parameters of the model were kept the same as previously published instantiations of the model, because these parameters were already optimized to simulate the appropriate receptive field sizes and other conceptually important features of the visual system (see Giese & Poggio, 2003). The only significant change we made to the model was to change the optic flow input at the front end of the motion pathway, which was estimated using the Horn–Schunck algorithm. We tested the model with 5,000 simulated trials of “bubblized” walkers for each condition. We collected the accuracy of the responses from the highest level of the form and motion pathways on each trial, and based on performance, the number of bubbles was adjusted with the

same staircase procedure as the behavioral experiment, resulting in 80% model accuracy. The threshold number of bubbles was estimated by averaging the number of bubbles across the last 500 trials, analogous to the behavioral data. Likewise, we used the same methods as the behavioral experiment to compute spatiotemporal classification movies. The resulting classification movies represent the diagnostic information that each pathway of the model uses to accurately discriminate walking direction. It should be noted that we are essentially modeling “experienced” observers. Previous human behavioral evidence shows that observers significantly improve their performance after as little as 20 trials (Hiris, Krebeck, Edmonds, & Stout, 2005; Jastorff et al., 2006), thus we anticipate that both the model and human observers, which are exposed to over 1,500 trials, would have acquired expert performance.

Results

Like the human observers, each pathway of the model was able to discriminate the leftward from rightward walker given sufficient information revealed by the bubbles. A summary of bubble thresholds is reported in Table 1. The model’s threshold estimates in the PL condition were comparable to the human observers. However neither pathway of the model performed as well as human observers for the SF condition, which suggests that observers may have been utilizing additional information in stick figures that was not captured by the model. The form pathway did show increased sensitivity to SFs as compared to the PL animations, which is likely due to the explicit form information that is lacking in PL sequences.

Classification movies derived from model responses in each pathway are shown in Figure 2. The classification movies from the motion pathway (Figure 2B) very closely resembled those from the human subjects, namely, that there was significant diagnostic information contained in the lower body of the figure, in particular the feet. A correlation analysis revealed a significant positive correlation, $p < 0.01$, between the behavioral and motion pathway classification movies for both stimulus conditions (Table 1). Just like the human observers, the motion pathway showed strong preference for the ankles as they cross and the fully extended front ankle as it changes direction in full stride. However, the motion model also used features in the crossing knees, while observers failed to use this cue.

For SFs, there was also a considerable correspondence between the behavioral data and motion pathway, particularly in the knees and ankles as they crossed the midline. The motion pathway also derived diagnostic information from the front-leading arm, when that arm was positioned out in front of the body. Thus, in addition to the key features in the feet used by both the model and the human observers, the model’s classification movies revealed

additional potent dynamic features that were largely untapped by human observers.

The classification movies derived from the form pathway (Figure 2C) also shared some features with those obtained from the human observers, but the correlation coefficients were not nearly as strong as the motion pathway (Table 1). In the PL condition, the form pathway relied almost exclusively on upper body structural features, particularly in the front-most arm. Human observers also appeared to use upper body cues in specific phases of the gait cycle and in conjunction with the head and shoulders. There was a significant negative correlation between the PL behavioral and the form pathway classification movies, indicating a systematic difference in strategies. This suggests that in some cases the diagnostic features for observers were unreliable, or even misleading, for the form pathway, and vice versa.

In the SF condition, the most diagnostic features for the form pathway were contained in the front-most arm and the leading leg. This was not a feature used by the human observers and suggests that overall the form pathway only captured a small subset of the human data.

Together, the behavioral and modeling results illustrate the balance between form- and motion-based strategies. Each trial of the bubble experiment revealed limited spatial and temporal stimulus information and observers adapted strategies to take advantage of the limited information available to perform the direction discrimination task. In this experiment, it is dynamic visual cues that appear to dominate and generalize across depictions of human actions.

Experiment 2

A study by Thirkettle et al. (2009) illustrated that observer strategies can change depending on the temporal duration of PL stimuli used in their psychophysical experiments. To determine whether the specific timing parameters used in our experiment had biased our findings toward motion features, we ran a second experiment that varied stimulus duration from very short (67 ms) to intermediate durations (133 and 266 ms). We reasoned that limiting exposure time to only a few frames on each trial severely limits motion information, while leaving form information relatively intact. We hypothesized that the classification movies for PL walkers would show a shift from motion-based strategies at longer exposure times (e.g., 467 ms in Experiment 1), to form-based strategies with very short exposure times.

Participants

Twenty-eight (21 females, 7 males) unpaid undergraduate students were recruited from the University of California, Irvine research participation pool. None of

these students participated in the first experiment, and no participants had prior experience with PL biological motion. Participants reported normal or corrected-to-normal vision and gave written consent in accordance with the University's IRB protocol.

Methods and analysis

We repeated all of the same methods and procedures from Experiment 1, with a few notable exceptions. First, subjects were randomly assigned to one of three conditions, each with different PL exposure durations: 67 ms (2 frames), 133 ms (4 frames), or 266 ms (8 frames). Ten subjects participated in the 133- and 266-ms conditions, each completing three blocks of 500 trials. Eight subjects participated in the 67-ms condition. Due to the shorter trial duration, we collected four blocks of 500 trials from seven subjects, and two blocks of 500 trials from one subject. In total, 15,000 trials were collected for each condition.

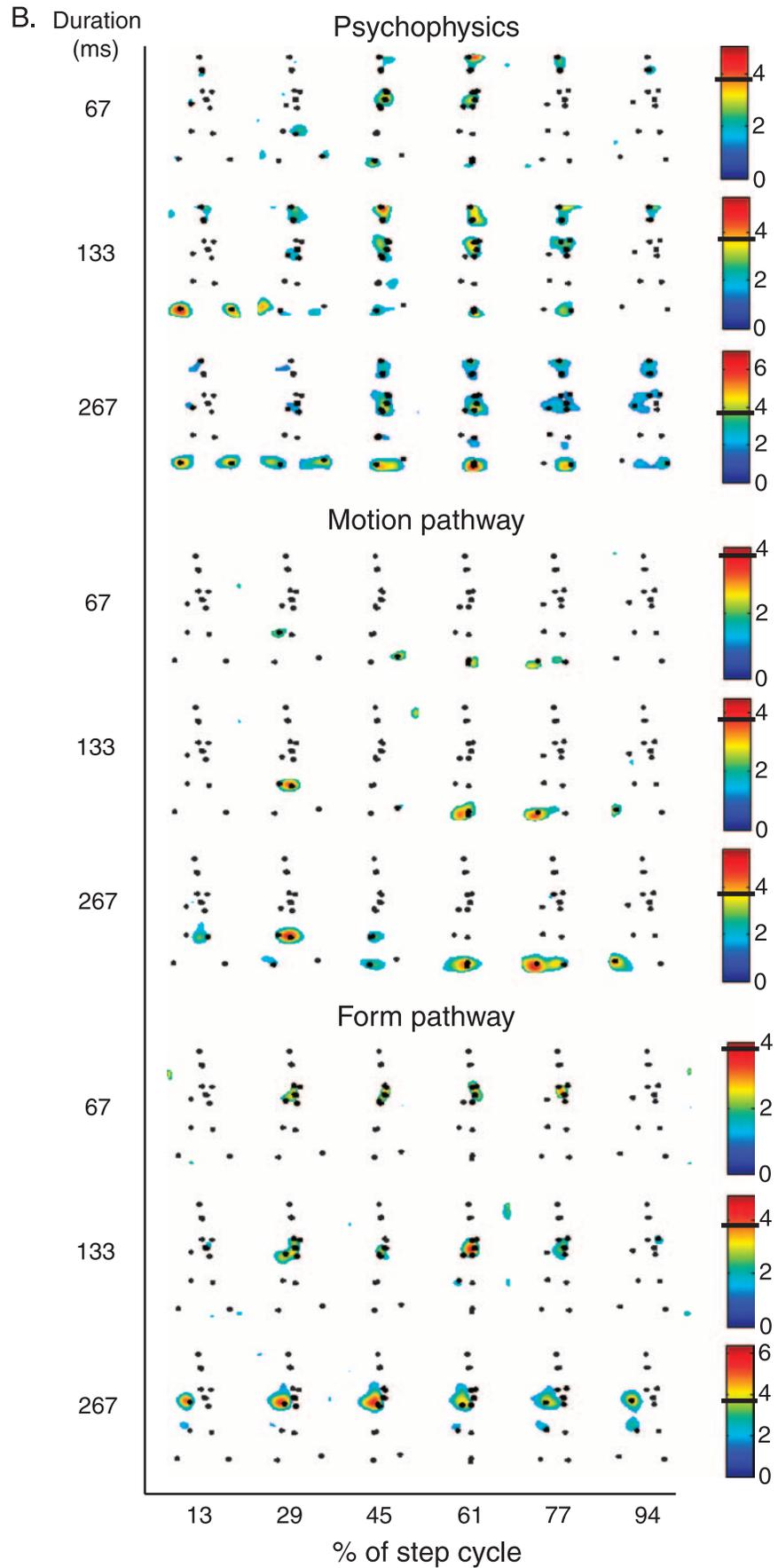
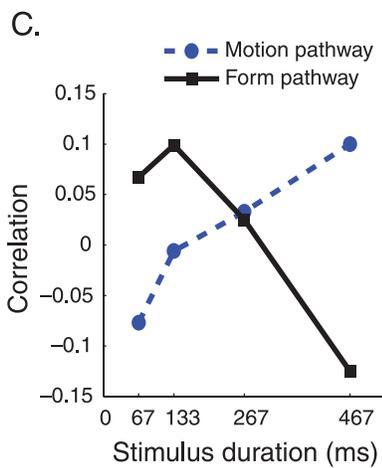
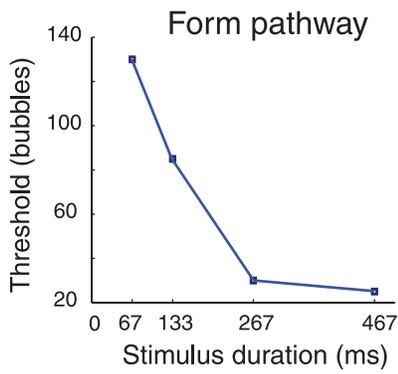
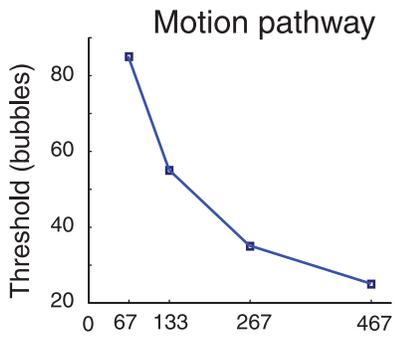
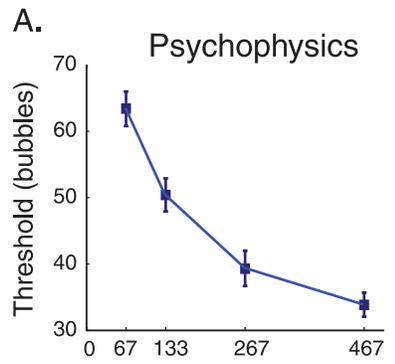
Similar to Experiment 1, the “bubblized” animation sequence onset between 33 and 233 ms after the start of the trial. The number of bubbles in the mask at staircase onset for the first block was initiated with 40, 35, and 30 bubbles in the 67-ms, 133-ms, and 266-ms conditions, respectively. The staircase procedure and analysis of classification movies was identical to Experiment 1.

We also ran model simulations with varying exposure times (67 ms, 133 ms, and 266 ms) corresponding to those used in the current experiment. The starting number of bubbles for the model simulations was determined by a pilot experiment that estimated the appropriate threshold for each condition and each pathway separately. These estimates were used as the starting number of bubbles for the simulations. The procedure for collecting and analyzing model data was otherwise the same as in Experiment 1.

Results

Bubble thresholds for each condition are plotted in Figure 3A. Sensitivity, as measured by the number of bubbles needed for threshold performance, increased quickly as sufficient spatiotemporal information became available. The data illustrate that sensitivity increased

Figure 3. Classification movies, thresholds, and correlation results from Experiment 2. (A) Bubble thresholds as a function of stimulus duration for human observers (psychophysics) and the model. Error bars indicate ± 1 standard error of the mean. (B) Classification movie results for all stimulus duration conditions, as derived from the human observers (psychophysics) and the model (motion and form pathways). Classification movies are visualized as described in Figure 2. (C) Correlation coefficients ($n = 185,964$ pixels) between behavioral and model classification movies for all duration conditions, including Experiment 1 (467 ms).



non-linearly as exposure time increased, and leveled off after about 266 ms. Such a non-linearity in the integration of information in biological motion is consistent with the initial reports of the minimum information needed to recognize biological motion (Johansson, 1973) and previous reports of duration thresholds (Thurman & Grossman, 2008).

The bubble thresholds derived from the computational model yielded a very similar pattern of results. Both the form and motion pathways required fewer bubbles with increasing exposure durations and showed a similar non-linear increase in sensitivity with an asymptote at a few hundred milliseconds.

Classification movies derived from behavioral responses and model responses in each pathway are shown in Figure 3B. Human observer classification movies over space and time were similar across the three durations, with a few key differences. At the shortest duration (67 ms or 2 frames), observers relied almost exclusively on upper body form cues, most notably the shoulders and head. In the shortest duration condition, this feature was most diagnostic at the point in the gait cycle when the limbs were aligned, but at longer exposure durations was diagnostic throughout the gait cycle. Thus, the head and shoulders appear to be a key form cue that is diagnostic of facing direction and readily available to human observers.

At the longer exposure durations, observers also use diagnostic features in the ankles. The strength of this cue increases with stimulus duration, but even with as little as 133 ms (4 frames) the features contained in the movement of the feet became reliably diagnostic.

The model performance at these varying exposure durations captured components of the human performance. The motion pathway of the model largely replicated the findings of the lower body, with the same reliance on the movement of the feet that strengthened with longer exposure durations. The form pathway relied exclusively on upper body features at all exposure durations with a pattern quite similar to observers.

The correlation analysis between the behavioral classification movies and each pathway illustrates the transition from form-based strategies with short exposures to motion-based strategies with longer exposure times (Figure 3C). There was a significant positive correlation ($p < 0.01$) between behavioral results and the form pathway at exposure times less than 267 ms. In contrast, there was a significant positive correlation ($p < 0.01$) between behavioral results and the motion pathway when the exposure time was greater than 267 ms. These data suggest that the transition from form- to motion-based strategies occurs between 200 and 300 ms.

Discussion

In the present series of experiments, we combined psychophysical and computational methods to investigate

the critical space–time features for biological motion perception. These experiments used the bubbles technique to restrict the information available to observers in the spatial and temporal domains randomly on each trial, and then assessed sensitivity to space–time features using reverse correlation. The result of this method is the diagnostic information that observers use to discriminate the walking direction of PL and SF biological motions.

The results from our classification movies revealed that human observers used a combination of form and motion cues, regardless of the type of visual presentation (PL or SF). In particular, observers relied heavily on upper body posture (head and shoulders) and lower body dynamics (the feet) when discriminating the facing direction of a PL or SF walker. These findings replicate previous work on the importance of upper body alignment on these discriminations (Lange et al., 2006) and the importance of dynamic information available in the feet (Casile & Giese, 2005; Mather, Radford, & West, 1992; Troje & Westhoff, 2006).

These interpretations are corroborated by results from model simulations, primarily from the performance of the computational motion pathway. The motion pathway discriminated walking direction by relying heavily on the lower body, tracking the feet and knees as they move in the gait cycle. Specifically, because the second layer of the model computed local opponent motion, the classification movies illustrated primarily those temporal intervals and locations that contained this mid-level motion feature. The motion pathway failed to reveal dynamic cues in the head and shoulders, which is not surprising given the relative immobility of these points. In contrast, the form pathway used largely different key form features than those used by the human observers at long exposure durations. The key features from the form pathway were isolated to the leading arm and upper leg. Thus, although task-related diagnostic form features exist in the PL and SF animations, human observers appear to be relatively insensitive to them.

In the second set of experiments, we manipulated the exposure time of the PL walker. Both the observers and the computational model showed the same non-linear decrease in threshold as exposure time increased. The observer classification movies, and the correlation analysis with the model's classification movies, revealed a gradual shift from using motion features at the longer durations to form-based features at the shortest durations. These data suggest that the transition point from using primarily form cues to motion cues occurs at around 200–300 ms. While observers did appear to use form cues at all durations, the motion features clearly dominated when the temporal duration was sufficiently long for the visual system to use the motion information.

Together, our findings characterize motion and form features that observers use to discriminate human walkers. Regions on the upper body, including the head and shoulders have minimal movement, and yet are diagnostic.

Observers used these form cues for discriminating both the SF and PL animations, suggesting that implied form information in biological motion is as potent as when the form is made explicit. In previous work, Lange et al. (2006) emphasized the importance of upper body posture as a form cue, arguing that the spatial displacement of the limbs from the vertical meridian is a diagnostic feature for discriminating walking direction, as performed by their template-matching model. The classification movies from our model appeared to pick up on this cue and additional postural cues, for example in the position of the front-most arms and legs. Yet observers were not particularly sensitive to those features of the walkers. We conclude that observers simply do not behave as ideal observers. This is a finding that was demonstrated in previous work (Gold, Tadin, Cook, & Blake, 2008) comparing human performance to an ideal observer and showing that humans are not particularly efficient with full-figure or PL biological motion.

At least two motion cues have been identified that facilitate biological motion perception: the relative opponent motion of the limbs and acceleration of the feet. Opponent motion occurs when the limbs cross, such as the relative motion of the feet or the swinging of the arms. Acceleration cues are prominent in the ankles, for instance, when the ankles reach the peak of their trajectory and change direction to head toward the midline. Artificial PL animations constructed from opponent motion features are often identified as biological (Casile & Giese, 2005), and PL walkers with acceleration cues removed are more difficult to perceive (Chang & Troje, 2009).

An additional consideration is the likely task dependency of these key features, as observers can adopt unique strategies as experimental conditions demand. For example, Thirkettle et al. (2009) recently demonstrated that while detection in motion-matched noise caused observers to adopt a motion-based strategy, discriminating very brief biological from scrambled animations without noise caused observers to adopt a form-based strategy. We likewise show here that changing the stimulus duration can affect observer strategies by encouraging form analysis for short durations and motion analysis for durations longer than 267 ms. Thus, it is imperative to consider task demands, and especially temporal duration parameters, when interpreting behavioral data from PL experiments.

The spatiotemporal bubbles technique as a means for identifying critical diagnostic information has a number of strengths and weaknesses. One disadvantage is the artificial nature of the experimental conditions, breaking up the biological kinematics into a number of space–time fragments that must be perceptually integrated. In the point-light condition, however, this could be considered a more sparse depiction of an already degraded animation. Moreover, because the bubbles technique is best suited to measuring explicit diagnostic parts of an image, as

opposed to reconstructing behavioral receptive fields (Gold, Murray, Bennett, & Sekuler, 2000), this technique is particularly well suited to recover local space–time features contained in individual point lights.

In considering the strengths and weaknesses of the bubbles technique in estimating key diagnostic features, Murray and Gold (2004) have argued that the bubbles method, in practice, may bias observers to use local features for discrimination that they would not otherwise if they had access to the global stimulus. It is important to note, however, that bubbles is a “self-calibrating” method, and that the staircase procedure would reveal any global or large-scale relational features that were *necessary* for discrimination (Gosselin & Schyns, 2001). In practice, this would be evidenced by an increase in the threshold number of bubbles for our tasks such that global structure would be reliably revealed. At the same time, the size of the bubble used in the experiment limits the minimum spatial extent of the diagnostic features, which in our case was approximately the size of a single point-light token. Considering this, our results clearly show that specific local space–time features (subsets of the overall structure) are *sufficient* for the discriminations in our task.

An advantage of the bubbles method, however, is overcoming the inherent limitations associated with adding external noise to the stimulus, as is used in dot masking techniques (Hiris, 2007). Although dot masks are a common means for testing biological motion sensitivity, biological motion is notoriously difficult to mask in noise (Hiris, Humphrey, & Stout, 2005) unless the animation is inverted (Pavlova & Sokolov, 2000) or presented in the periphery (Thompson, Hansen, Hess, & Troje, 2007). The most effective masks are those that are closely matched to the local space–time features of the actor, such as small triads of dots or scrambled walker masks (Bertenthal & Pinto, 1994). Although ideal for controlling psychophysical performance, matching noise masks have the unfortunate effect of forcing the observer to perceptually organize the animation into groups of signal and noise, a step that is inherently intertwined with the discrimination task. Hence, there may be ambiguity as to whether the measurement sensitivity is driven by key features in the biological motion, or by the figure–ground segmentation itself (Thompson et al., 2007).

Because the bubbles do not introduce clutter or noise, we essentially eliminate the process of segmentation. Instead, Gaussian apertures reveal portions of the animation, similar to viewing a person walking through a forest with occluding leaves in the visual field, for instance. Importantly, the bubble mask can be suitably applied to other renditions of biological motion like SFs or even fully illuminated natural images, which would otherwise be non-trivial to obscure with an equivalent mask.

The current technique is similar in spirit to the correlation map technique introduced by Lu and Liu (2006). In an elegant study, the researchers used reverse

correlation to compute dynamic classification movies using white noise masks in a forward/backward walking discrimination task. Their analysis revealed classification movies showing that each individual point light, on average, contributes equally to the discrimination task, supporting the hypothesis of global analysis of biological motion.

The discrepancy between those results and our current results may be attributed to a number of possible factors. Most critically, in our experiment, observers viewed a short segment of the overall gait cycle on each trial, which served to reveal the temporal dynamics in the diagnostic space–time features. In contrast, Lu and Liu (2006) displayed a full gait cycle on each trial and their analysis would not capture fluctuations in the behavioral significance of specific postures or individual point lights over time (Thirkettle et al., 2009; Thurman & Grossman, 2008).

A second consideration is the inherent strengths and weaknesses of bubbles and white noise analyses in revealing local and global features, respectively. Because the bubbles technique may serve to emphasize local features, our experiments were more closely designed to reveal diagnostic local space–time features. A weakness, however, is that our experimental design may also have served to encourage observers to adopt a more local strategy, while the white noise mask may encourage a global strategy (Murray & Gold, 2004). This is an important consideration for comparing the two experiments, and thus we interpret our findings as complementary to, and not contradictory of, those findings of Lu and Liu (2006).

Lastly, differences in task demands and stimulus generation may have contributed to the differing results. Our subjects participated in a left/right walking direction discrimination task on motion-captured sequences, while subjects in Lu and Liu's (2006) study completed a forward/backward discrimination task on Poser-generated animations. A recent study illustrated that local motion cues in the feet are important for discriminating walking direction (Saunders, Suchan, & Troje, 2009), particularly with PL displays derived from motion capture data and that artificial PL walkers generated from the Cutting algorithm (Cutting, Proffitt, & Kozlowski, 1978) differ in perceptually relevant ways from motion-captured walkers. Thus, these simple stimulus differences may have also played a role in the differing results.

Overall, the space–time bubbles technique renders classification movies with sufficient spatial and temporal resolutions for identifying key features in biological motion. We have also shown this to be an apt tool for illustrating the stimulus features that modulate responses of a computational model. Although in practice it is sometimes unclear exactly what features are modulating model responses, this method can be a useful tool for testing and understanding models of perception.

Acknowledgments

This work was supported by a grant from the National Science Foundation (BCS0748314) to E. Grossman. M. Giese was supported by the Deutsche Forschungsgemeinschaft, EC Project COBOL, and the Hermann Lilly Schilling Foundation.

Commercial relationships: none.

Corresponding author: Steven M. Thurman.

Email: sthurman@uci.edu.

Address: 3151 Social Science Plaza, Irvine, California, USA.

References

- Aaen-Stockdale, C., Thompson, B., Hess, R. F., & Troje, N. F. (2008). Biological motion perception is cue-invariant. *Journal of Vision*, 8(8):6, 1–11, <http://www.journalofvision.org/content/8/8/6>, doi:10.1167/8.8.6. [PubMed] [Article]
- Ahlström, V., Blake, R., & Ahlström, U. (1997). Perception of biological motion. *Perception*, 26, 1539–1548.
- Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14, 105–126.
- Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial actors in gait perception that influence gender recognition. *Perception & Psychophysics*, 23, 145–152.
- Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 5661–5663.
- Bertenthal, B., & Pinto, J. (1994). Global processing of biological motion. *Psychological Science*, 5, 221–225.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*, 5(4):6, 348–360, <http://www.journalofvision.org/content/5/4/6>, doi:10.1167/5.4.6. [PubMed] [Article]
- Chang, D. H., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision*, 9(1):19, 1–17, <http://www.journalofvision.org/content/9/1/19>, doi:10.1167/9.1.19. [PubMed] [Article]
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5(9):1,

- 659–667, <http://www.journalofvision.org/content/5/9/1>, doi:10.1167/5.9.1. [PubMed] [Article]
- Cutting, J. E., & Kozlowski, L. T. (1977). Recognition of friends by their walk. *Bulletin of the Psychonomic Society*, *9*, 353–356.
- Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 357–372.
- Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*, 15–22.
- Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*, 727–738.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*, 482–498.
- Fiset, D., Blais, C., Arguin, M., Tadros, K., Ethier-Majcher, C., Bub, D., et al. (2008). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology*, *26*, 23–35.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, *76*, 2718–2739.
- Garcia, J. O., & Grossman, E. D. (2008). Necessary but not sufficient: Motion perception is required for perceiving biological motion. *Vision Research*, *48*, 1144–1149.
- Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179–192.
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, *10*, 663–666.
- Gold, J. M., Tadin, D., Cook, S. C., & Blake, R. (2008). The efficiency of biological motion perception. *Perception & Psychophysics*, *70*, 88–95.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, *15*, 97–112.
- Gosselin, F., & Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271.
- Grossman, E. D., & Blake, R. (1999). Perception of coherent motion, biological motion and form-from-motion under dim-light conditions. *Vision Research*, *39*, 3721–3727.
- Hegde, J., & van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, *20*, 1–6.
- Hiris, E. (2007). Detection of biological and nonbiological motion. *Journal of Vision*, *7*(12):4, 1–16, <http://www.journalofvision.org/content/7/12/4>, doi:10.1167/7.12.4. [PubMed] [Article]
- Hiris, E., Humphrey, D., & Stout, A. (2005). Temporal properties in masking biological motion. *Perception & Psychophysics*, *67*, 435–443.
- Hiris, E., Krebeck, A., Edmonds, J., & Stout, A. (2005). What learning to see arbitrary motion tells us about biological motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1096–1106.
- Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, *160*, 106–154.
- Jastorff, J., Kourtzi, Z., & Giese, M. A. (2006). Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision*, *6*(8):3, 791–804, <http://www.journalofvision.org/content/6/8/3>, doi:10.1167/6.8.3. [PubMed] [Article]
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 195–204.
- Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from dynamic point-light display. *Perception & Psychophysics*, *21*, 575–580.
- Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision*, *6*(8):6, 836–849, <http://www.journalofvision.org/content/6/8/6>, doi:10.1167/6.8.6. [PubMed] [Article]
- Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *Journal of Neuroscience*, *26*, 2894–2906.
- Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE Transactions Pattern Analysis and Machine Intelligence*, *28*, 802–817.
- Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object vision. *Annual Review of Neuroscience*, *19*, 577–621.
- Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision*,

- 6(4):12, 475–483, <http://www.journalofvision.org/content/6/4/12>, doi:10.1167/6.4.12. [PubMed] [Article]
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop*, 121–130.
- Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shape. *Proceedings of the Royal Society of London B*, 214, 501–524.
- Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London B: Biological Sciences*, 249, 149–155.
- Murray, R. F., & Gold, J. M. (2004). Troubles with bubbles. *Vision Research*, 44, 461–470.
- O'Rourke, J., & Badler, N. (1980). Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2, 522–536.
- Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, 62, 889–898.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82, B51–B61.
- Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 358, 435–445.
- Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, 14, 302–308.
- Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. *Journal of Vision*, 9(6):15, 1–32, <http://www.journalofvision.org/content/9/6/15>, doi:10.1167/9.6.15. [PubMed] [Article]
- Saunders, D. R., Suchan, J., & Troje, N. F. (2009). Off on the wrong foot: Local features in biological motion. *Perception*, 38, 522–532.
- Smith, A. T., & Snowden, R. J. (1994). *Visual detection of motion*. London: Academic Press.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19, 109–139.
- Tanaka, K., Fukuda, Y., & Saito, H. (1989). Analysis of motion of the visual field by direction expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62, 626–641.
- Thirkettle, M., Benton, C. P., & Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *Journal of Vision*, 9(3):28, 1–11, <http://www.journalofvision.org/content/9/3/28>, doi:10.1167/9.3.28. [PubMed] [Article]
- Thompson, B., Hansen, B. C., Hess, R. F., & Troje, N. F. (2007). Peripheral vision: Good for biological motion, bad for signal noise segregation? *Journal of Vision*, 7(10):12, 1–7, <http://www.journalofvision.org/content/7/10/12>, doi:10.1167/7.10.12. [PubMed] [Article]
- Thurman, S. M., & Grossman, E. D. (2008). Temporal “bubbles” reveal key features for point-light biological motion perception. *Journal of Vision*, 8(3):28, 1–11, <http://www.journalofvision.org/content/8/3/28>, doi:10.1167/8.3.28. [PubMed] [Article]
- Todd, J. T. (1983). Perception of gait. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 31–42.
- Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2(5):2, 371–387, <http://www.journalofvision.org/content/2/5/2>, doi:10.1167/2.5.2. [PubMed] [Article]
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector”? *Current Biology*, 16, 821–824.
- Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics*, 67, 667–675.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Vangeneugden, J., Pollick, F., & Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex*, 19, 593–611.
- Vinette, C., Gosselin, F., & Schyns, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science: A Multidisciplinary Journal*, 28, 289–301.
- Wang, J. Y. A., & Adelson, E. H. (1994). Representing moving images with layers. *IEEE Transactions on Image Processing*, 3, 625–638.