CrossMark

Revisiting the importance of common body motion in human action perception

Steven M. Thurman[1] · Hongjing Lu[1,2]

Abstract Human actions are complex dynamic stimuli comprised of two principle motion components: 1) common body motion, which represents the translation of the body when a person moves through space, and 2) relative limb movements, resulting from articulation of limbs after factoring out common body motion. Historically, most research in biological motion has focused primarily on relative limb movements while discounting the role of common body motion in human action perception. The current study examined the relative contribution of posture change resulting from relative limb movements and translation of body position resulting from common body motion in discriminating human walking versus running actions. We found that faster translation speeds of common body motion evoked significantly more responses consistent with running when discriminating ambiguous actions morphed between walking and running. Furthermore, this influence was systematically modulated by the uncertainty associated with intrinsic cues as determined by the degree of limited-lifetime spatial sampling. The contribution of common body motion increased monotonically as the reliability of inferring posture changes on the basis of intrinsic cues decreased. These results highlight the importance of translational body movements and their interaction with posture change as a result of relative limb movements in discriminating human actions when visual input information is sparse and noisy.

Gunnar Johansson's seminal article in 1973 introduced a novel tool for studying biological motion perception using point-lights or small markers to represent the joint movements of humans in action (Johansson, 1973). This work provided a foundation for an active field of research in biological motion that has now spanned more than 40 years (Blake & Shiffrar, 2007). Johansson discussed biological motion perception in the context of his theory of vector analysis, which highlighted the importance of local image motion and reference frames for dynamic event perception. According to this theory, moving objects are composed of two principle motion components: 1) common body motion which represents translation of the entire object within the environmental reference frame, and 2) relative motion of objects parts within the object-centered reference frame after removing or factoring out common body motion. Johansson proposed that a primary role for common body motion was to provide a means for grouping the intrinsic cues in order to analyze their relative movements within a common reference frame. In the literature of biological motion perception, common body motion also is termed as extrinsic motion and relative movements of object parts as intrinsic motion.

Consequently, relative motion of object parts was theorized to play a more prominent role in recognizing and distinguishing among different types of dynamic events. For example, Johansson made the astute observation that

✉ Steven M. Thurman
sthurman@ucla.edu

✉ Hongjing Lu
hongjing@ucla.edu

[1] Department of Psychology, University of California, Los Angeles, CA 90095, USA

[2] Department of Statistics, University of California, Los Angeles, CA 90095, USA

removing common body motion from a point-light walker stimulus (making it appear to walk on a treadmill) had no detrimental impact on the ability of observers to immediately perceive the stimulus as a human walker (Johansson, 1973). Moreover, this finding led to a rather strong claim that "the proximal motion patterns…carry all the essential information needed for immediate visual identification of such human motions" (pp. 210) in which "proximal" motion corresponds to relative limb motion defined in the present paper. No doubt, this idea has had a profound impact on the methodology and scope of research within the field of biological motion over the last four decades, with a majority studies showing a distinct focus on relative limb movements, while discounting the potential relevance of common body motion. Some studies have included common body motion in point-light displays (Murphy, Brady, Fitzgerald, & Troje, 2009; Oram & Perrett, 1996; Oram & Perrett, 1994), but few have examined systematically the interaction between common and relative aspects of biological motion (but see a recent publication by Masselink & Lappe, 2015). For work to be relevant to this issue, the minimal requirement is that both aspects of biological motion should be present and manipulated in experimental displays.

The present paper aims to address this issue by manipulating the reliability of visual cues critical for relative limb movements and the validity of common body motion in the display. Imagine a person moving through an environment with dense foliage, or perhaps partially occluded by other objects or people. At each moment, you might only get sparse bits of information about the shape of the person, the dynamics of their limbs, and their overall position in relation to the environment. If the foliage or occluding objects were very sparse, you might expect large glimpses of information about body postures and individual limb movements and therefore would be able to use these intrinsic body cues due to their high reliability and potential usefulness. However, under densely occluded conditions, this local information might prove too sparse and unreliable, in which case it might be advantageous to make use of information based on the relative speed of the entire body, which may be reliably estimated from the common motion of the disjointed patches over a relatively longer range of duration and larger spatial regions. Hence, under conditions with sparse or unreliable information about relative limb movements, observers might be well served to incorporate common body motion cues into their judgments.

The current study was designed to emulate this naturalistic example but under well-controlled laboratory conditions. To make the action perception task challenging, the experiment included ambiguous action stimuli that were systematically morphed between the prototypical human actions of walking and running. The stimulus moved horizontally across the screen at different translational speeds as determined by the speed of common body motion, which ranged from 3 MPH (consistent with the speed of human walking) to 6 MPH (consistent with a moderate running speed). By varying translation speed and morph level, we created situations of cue conflict between extrinsic and intrinsic cues, for instance where body samples would represent the postures of a walking-like action but where translation speed would be faster and therefore consistent with a running-like action. In addition, the point-light stimuli consisted of a number of limited-lifetime dots sampled along the underlying skeleton to manipulate the uncertainty associated with the inference of posture change over time on the basis of intrinsic limb movements (Beintema, Georg, & Lappe, 2006; Beintema & Lappe, 2002; Thurman & Lu, 2014a). As a result, the three factors involved in our design allow us to examine the interaction between common body motion and relative limb movements to the judgment of discriminating different actions.

Methods

Participants

Sixteen UCLA undergraduate students (mean age = 22.3, 9 females) with normal or corrected vision enrolled the study to complete course credit for their participation. Participants gave informed consent as approved by the UCLA Institutional Review Board.

Stimuli and procedure

Motion capture data of human activities were retrieved from CMU motion capture database (http://mocap.cs.cmu.edu) and displayed using the Biomotion Toolbox (van Boxtel & Lu, 2013b). The actions included a human figure walking or running in the sagittal (profile) view. The horizontal translation component was removed on a frame-by-frame basis, and gait cycle duration was equalized between the walking and running stimuli (1.18 s) to facilitate morphing between the actions across an equivalent number of underlying postures. Equalizing gait cycle also helped to ensure that observers would rely on information provided by distinct body postures and would prevent them from adopting a strategy of merely counting the number steps taken by the actor or relying on perceived duration of the gait cycle. To ensure that the manipulation of gait cycle duration did not reduce the perceived naturalness of human walking and running actions, an additional online experiment recruited 52 subjects through Amazon Mturk. Two point-light actions, walking and running, were rendered with their natural gait cycle duration or with equalized gait cycle duration. These action stimuli were translated across the screen at four speeds, 1.5, 3.0, 6.0, and 9.0

MPH. Participants were asked to rate the action stimuli in terms of "how natural the person appeared to move across the screen," using a 5-point rating scale. We found no significant difference in the pattern of results for the equalized gait stimuli compared with the normal (e.g., natural) gait stimuli.

Next, we employed a motion-morphing algorithm to generate stimuli that were morphed between these prototypical actions (Giese & Lappe, 2002). A morph value of 0 was defined as a true runner and a value of 1 defined as a true walker. Hence, a morph value of 0.2 would represent a combination of 80 % running with 20 % walking. We varied the morph value systematically across 7 equidistant levels ranging from 0.125 to 0.875 (in steps of 0.125) to create point-light actions with varying levels of ambiguity in terms of whether the stimulus would appear more like a walker or runner (Fig. 1).

Next, the morphed actions were resampled to create sparse animations using the 1-frame limited-lifetime technique (Beintema et al., 2006; Beintema & Lappe, 2002). For each trial, we specified a number of spatial samples that would be randomly drawn from the underlying body posture on each frame of the sequence (Fig. 1). Sampling was constrained such that at least one randomly sampled point was placed in each limb before re-sampling the limbs when the number of points was greater than 8 (e.g., total number of limbs). Hence, the locations of these samples varied randomly from frame to frame, which severely disrupted the usefulness of local motion cues and forced the observer to rely more strongly on

form-based postural cues (Lange & Lappe, 2006; Theusner, de Lussanet, & Lappe, 2014; Thurman & Lu, 2014a). In the current experiment, we varied the number of sampling points across three levels (8, 12, and 16 points/frame), which provided a means to manipulate the sparseness, or reliability, of intrinsic body cues (Fig. 1). It should be noted that the reliability of information provided by common body motion is much less affected by this manipulation since every point has the same translation speed and direction. In other words, common motion represents a single vector quantity that can be reliably estimated from just a single point, whereas extracting information about body postures is a more complex computation that benefits from more sampling information.

On each trial, a morphed action was presented for 1 sec duration, with a randomized inter-trial interval between 1 and 2 sec. Stimuli comprised 60 frames presented at 60 Hz (17 ms per frame). Common body motion cues were introduced by translating the figure rightward across the screen at a variable speed (3.2, 4.8, or 6.4 deg/sec). Assuming a realistic human height of 70 inches, this corresponds to roughly 3, 4.5, and 6 miles per hour (MPH). These values are well-matched the naturalistic translation speed of the motion captured sequences from which the morphed stimuli were derived (3.2 and 7.6 deg/sec for walking and running, respectively). These speeds also are consistent with biophysical measurements, where the transitional speed between human walking and running actions ranges from 3 MPH to 5.7 MPH, with a mean of 4.3 MPH (Thorstensson & Roberthson, 1987). There is a
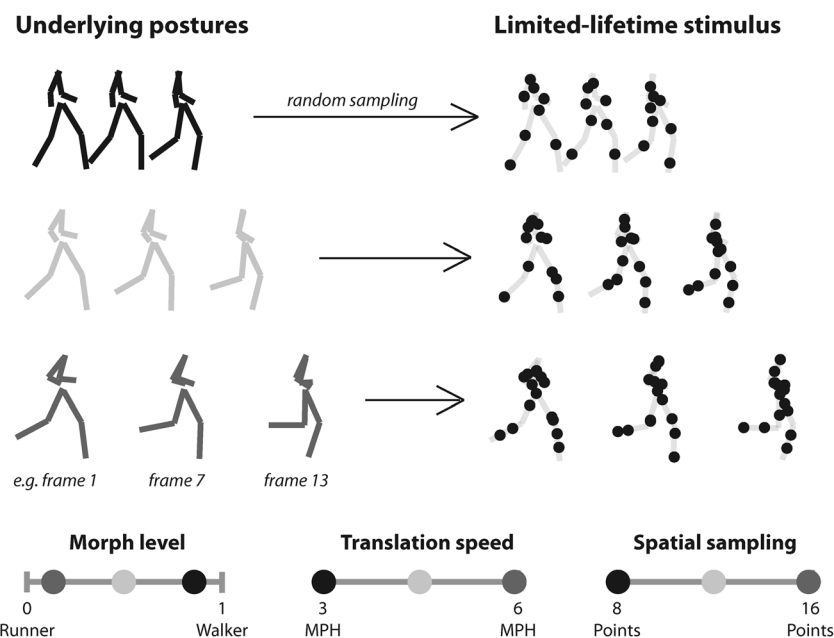

Fig. 1 Various types of stimulus manipulations. The morph level determined the degree to which the stimulus appeared to be walking vs. running within the range of 0 (true runner) to 1 (true walker). For instance, the figures shown in black, light gray, and dark gray had morph levels of 0.875, 0.5, and 0.125, respectively. The translation speed varied from 3.0 to 6.0 miles/hour (MPH), as illustrated by the horizontal spatial shift in individual postures over time. Spatial sampling determined the quantity of postural information provided by intrinsic cues and varied across three levels, 8, 12, and 16 points/frame

direct relationship between the height of a person and the speed of locomotion, such that speed is invariant to viewing distance if the height of the actor is known or easily estimated. Hence, perceptual judgments in this task on the basis of translation speed were likely not confounded by ambiguity or uncertainty about viewing distance.

We randomized the starting location of the figure on the screen within a rectangular region (5.6 by 5.6 deg) offset from the center of the screen to the left by 5.6 deg. The purpose of randomizing starting position was to minimize the use of absolute positional cues for action discrimination by introducing uncertainty to the starting and ending position of the figure on each trial.

The experiment and stimuli were programmed using Matlab (Mathworks, Inc) and the psychophysics toolbox (Brainard, 1997; Pelli, 1997) running on a Dell PC and presented on a calibrated CRT monitor (60 Hz, background luminance 16.2 c/m^2). Participants were seated in a darkened room with a viewing distance of 35 cm, and a chinrest was provided to minimize head movements during the experiment. The point-light figures were represented by high contrast black dots (diameter 0.5 deg) on a grey background and subtended 8 deg vertically.

Participants performed a forced choice discrimination task, indicating whether the moving stimulus appeared more like the action of a human walking or running by pressing a key on the keyboard. The experiment had a 7 x 3 x 3 within-subjects design, with 7 levels of morphing, 3 levels of spatial sampling, and 3 levels of translation speed. All trial types were balanced and randomly mixed within four blocks of 126 trials, totaling 504 trials and 8 trials per condition. The experiment was completed within a 1-hour session.

Analysis

We computed the proportion of running responses for each condition to fit individual subject data with cumulative Gaussian psychometric function, producing an estimate of the slope and bias for each experimental condition. The inverse of the slope parameter reflects the sensitivity of discriminating between walkers and runners as a function of morph level, where a steeper slope indicates higher sensitivity and therefore higher reliance on intrinsic cues for perceptual discrimination. The bias parameter reflects the point of subjective equality (PSE) or the morph weight that would produce 50 % walker/runner responses, which is visualized as a horizontal shift in the psychometric curve. The change in PSE (fast – slow translation) was taken to represent the influence of common body motion cues to responses in the action discrimination task. That is, if faster common body motion biased perception toward running, we would expect the

psychometric curve to shift to the right and the change in PSE to be greater than zero.

Results

Figure 2 depicts the group average probability of a running response as a function of morph level with the corresponding psychometric curve fits. To determine the influence of spatial sampling and the speed of translation on discrimination sensitivity across subjects, we performed a 3 x 3 repeated measures analysis of variance (ANOVA) on sensitivity (1/slope) in recognizing between a walker and a runner (Fig. 3a). A significant main effect of the number of sampling points, $F(2,30) = 21.10$, $p < 0.001$, was revealed to show that more spatial samples enhanced the reliance on intrinsic body cues to yield greater sensitivity in discriminating walking versus running. The main effect of translation speed was not significant, $F(2, 30) = 0.16$, $p = 0.77$, nor was the interaction between the two factors, $F(4,60) = 0.49$, $p = 0.64$, suggesting the lack of influence by the speed of common body motion to action discrimination sensitivity.

Next, we examined the influence of spatial sampling and translation speed on the bias to perceive a particular action, measured as the PSE. The mean PSE across all conditions was significantly higher than 0.5 (mean = 0.53, $t(8) = 16.73$, $p < 0.001$), indicating a slight bias to perceive morphed actions as running. This may be due to a higher distinctiveness for running postures or may indicate a more narrowly tuned representation for walking compared to the running action. A 3 x 3 repeated measures ANOVA on the PSE revealed a significant main effect of translation speed, $F(2,30) = 17.92$, $p < 0.001$, due to an increase in PSE for stimuli with faster translation (Fig. 3b). The main effect of spatial sampling was not significant, $F(2,30) = 0.25$, $p = 0.66$. Importantly, we found the significant interaction between translation speed of body motion and spatial sampling that controls the reliability of cues resulting from relative limb movements, $F(4,60) = 2.91$, $p = 0.029$. The interaction effect was due to an increasing impact from the translation speed when cue reliability from relative limb movements was low due to the reduction of the number of spatial samples. Together, these results demonstrate a clear trade-off between the contribution of relative limb movement cues and common body motion cues to discrimination performance, as illustrated in Fig. 3c. Faster translation speeds tended to induce more runner responses, but this influence was systematically modulated by the degree of spatial sampling, or in other words, the reliability of visual cues provided by relative limb movements.
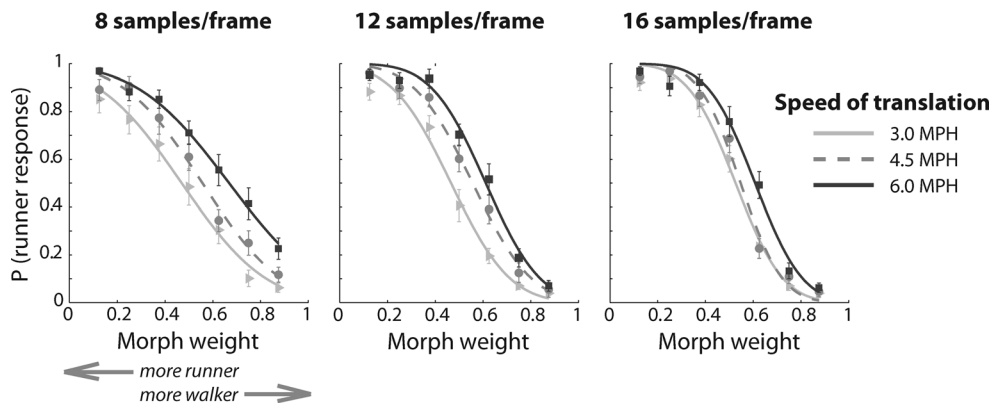
Fig. 2 Mean group results (n = 16) of the walking vs. running action discrimination task where the probability of a runner response is plotted as a function of morph weight. Group data shown were fit with cumulative Gaussian psychometric functions for illustration purposes. The horizontal curve shift between the fastest and slowest conditions

represents the change in point of subjective equality (PSE), and therefore the influence of common body motion cues to the action discrimination task. The slopes of the curves represent discrimination sensitivity, where 1/slope reflects the reliability of discriminations on the basis of intrinsic cues. Error bars represent SEM

Discussion

Previous research has shown that observers can accurately perform action discrimination based on intrinsic limb movements alone (Dittrich, 1993; Giese & Lappe, 2002; van Boxtel & Lu, 2013a). However, when intrinsic cues are completely removed, for instance, by representing the entire human figure as a single point or blob moving through space, observers are still capable of distinguishing several types of complex human interactions (McAleer, Kay, Pollick, & Rutherford, 2011; McAleer & Pollick, 2008). Even though observers *can* discount common body motion and perform many action recognition tasks, a key question regards the extent to which common body motion, as represented by the change of body position over time, should contribute to perceptual decisions in discriminating two actions.

The results of our study provide compelling evidence to highlight the interactions between common and relative motion components in biological motion perception. Consistent with previous findings, we found that translation speed impacted action discrimination performance. Critically, the strength of this influence was significantly modulated by the sparseness, or reliability, of intrinsic body form cues determined by the number of spatial elements sampled along the human figure. Of note, increasing the number of points beyond 8 would cause some limbs to be sampled with more than one element, which would likely facilitate the extraction of limb orientation information in addition to positional signals. Previous research has shown limb orientation to be a very potent visual cue for body form analysis (Lu, 2010; Thurman & Lu, 2013; Thurman & Lu, 2014a; Vangeneugden, Peelen, Tadin, & Battelli, 2014), and this also may have
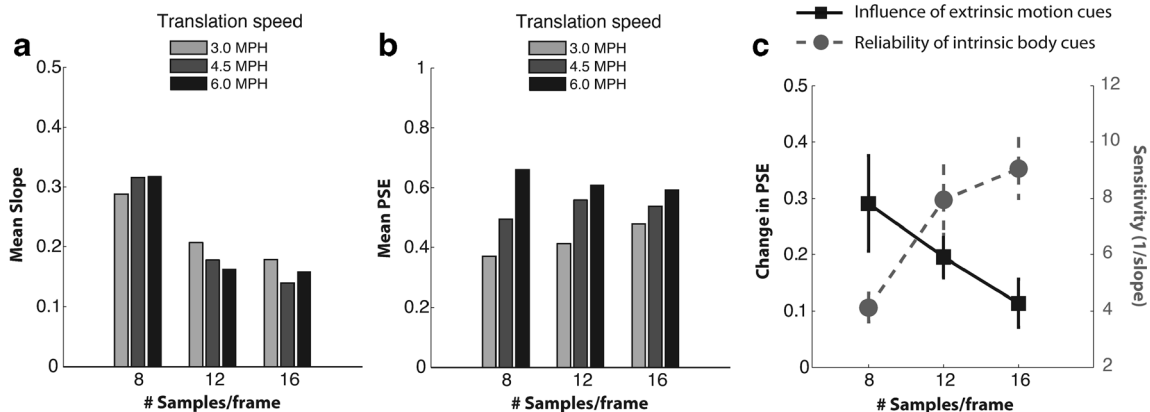

Fig. 3 Group results (n = 16) showing **a**) the mean slope parameter, and **b**) the mean bias parameter of the psychometric function (e.g., PSE) for each of the 9 conditions tested. **c**) The influence of extrinsic cues (shift in PSE for fast minus slow translation speeds) as a function of spatial sampling (left ordinate axis). Overlaid on the same graph, the reliability

of intrinsic cues reflecting the reliance on intrinsic body information (estimated as 1/slope parameter) as a function of spatial sampling (right ordinate axis). Notice the interaction between these two measurements, where the influence of extrinsic cues increases as the reliability of intrinsic cues decreases. Error bars represent SEM

contributed to the increased reliance on intrinsic body cues with increased spatial sampling.

The observed trade-off between intrinsic cue reliability and the influence of common body motion is illustrated graphically in Fig. 3c. These results are generally consistent with a Bayesian account that humans discriminate actions by taking into account information provided by multiple types of cues, and weighting the evidence provided by each source of information according to cue reliability (Thurman and Lu, 2014). Our findings are consistent with a recent study on simple object motion (Kwon, Tadin, & Knill, 2015) and highlight the importance of studying the interaction between common motion and relative component motion for dynamic stimuli. The convergence of these two studies, one from biological motion and another from nonbiological object motion, shows that both components should be considered together in order to gain the insights into underlying mechanisms.

Historically, most studies in the field of biological motion perception have focused on examining the relevance of limb movements and body postures for human action perception after discounting global body movements. This is evidenced by the numerous studies that employ the so-called "treadmill walker," in which global translation is intentionally removed in the experiment so that the figure will walk in place as if on a treadmill. However, in the present study, we show that humans are tuned in to common body motion information and incorporate it naturally with intrinsic body cues in the recognition of human actions. This avenue of research offers a distinctive perspective on biological action perception, with a renewed emphasis on the importance of common body motion cues and their interaction with action specific information provided by local body cues (Thurman & Lu, 2013, 2014b). To conclude, we have shown that human actions are more than a collection of body postures or joint movements within a static reference frame. Instead, the manner in which the entire body moves through the environment is shown to play a relevant and meaningful role in the representation and discrimination of human actions.

References

Beintema, J. A, Georg, K., & Lappe, M. (2006). Perception of biological motion from limited-lifetime stimuli. *Perception & Psychophysics*, *68*(4), 613–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16933426

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*(8), 5661–5663. doi:10.1073/pnas.082483699

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Perception*. doi:10.1146/annurev.psych.57.102904.190152

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*(1), 15–22.

Giese, M. A., & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Research, 42*(15), 1847–1858.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211. doi:10.3758/BF03212378

Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences, 112*(26), 8142–8147. doi:10.1073/pnas.1500361112

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *The Journal of Neuroscience, 26*(11), 2894–2906. doi:10.1523/JNEUROSCI.4915-05.2006

Lu, H. (2010). Structural processing in biological motion perception. *Journal of Vision, 10*(12), 13. doi:10.1167/10.12.13

Masselink, J., & Lappe, M. (2015). Translation and articulation in biological motion perception. *Journal of Vision, 15*(11), 10. doi:10.1167/15.11.10

McAleer, P., Kay, J. W., Pollick, F. E., & Rutherford, M. D. (2011). Intention perception in high functioning people with autism spectrum disorders using animacy displays derived from human actions. *Journal of Autism and Developmental Disorders, 41*(8), 1053–1063.

McAleer, P., & Pollick, F. E. (2008). Understanding intention from minimal displays of human activity. *Behavior Research Methods, 40*(3), 830–839.

Murphy, P., Brady, N., Fitzgerald, M., & Troje, N. F. (2009). No evidence for impaired perception of biological motion in adults with autistic spectrum disorders. *Neuropsychologia, 47*(14), 3225–3235.

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "Biological Motion" stimuli. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.1994.6.2.99

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, *76*(1), 109–129. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8836213

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. Retrieved from http://www.ingentaconnect.com/content/vsp/spv/1997/00000010/00000004/art00016

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. *The Journal of Neuroscience, 34*(3), 909–921. doi:10.1523/JNEUROSCI.2900-13.2014

Thorstensson, A., & Roberthson, H. (1987). Adaptations to changing speed in human locomotion: Speed of transition between walking and running. *Acta Physiologica Scandinavica, 131*(2), 211–214.

Thurman, S. M., & Lu, H. (2013a). Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space. *Journal of Vision, 13*(2), 8–8. doi:10.1167/13.2.8

Thurman, S. M., & Lu, H. (2013b). Physical and biological constraints govern perceived animacy of scrambled human forms. *Psychological Science, 24*(7), 1133–1141. doi:10.1177/0956797612467212

Thurman, S. M., & Lu, H. (2014a). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience, 8*, 91. doi:10.3389/fnhum.2014.00091

Thurman, S. M., & Lu, H. (2014b). Perception of social interactions for spatially scrambled biological motion. *PLoS ONE, 9*(11), e112539. doi:10.1371/journal.pone.0112539

Van Boxtel, J. J. A, & Lu, H. (2013a). Impaired global, and compensatory local, biological motion processing in people with high levels of autistic traits. *Frontiers in Psychology, 4*(April), 209. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3632794&tool=pmcentrez&rendertype=abstract

Van Boxtel, J. J. A., & Lu, H. (2013b). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *Journal of Vision , 13* (12). http://doi.org/10.1167/13.12.7

Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *The Journal of Neuroscience, 34*(2), 574–585. doi:10.1523/JNEUROSCI.4032-13.2014