Translation and articulation in biological motion perception

Jana Masselink

Institute of Psychology, University of Muenster, Münster, Germany

Markus Lappe

Institute of Psychology & Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Münster, Germany

Recent models of biological motion processing focus on the articulational aspect of human walking investigated by point-light figures walking in place. However, in real human walking, the change in the position of the limbs relative to each other (referred to as articulation) results in a change of body location in space over time (referred to as translation). In order to examine the role of this translational component on the perception of biological motion we designed three psychophysical experiments of facing (leftward/rightward) and articulation discrimination (forward/backward and leftward/rightward) of a point-light walker viewed from the side, varying translation direction (relative to articulation direction), the amount of local image motion, and trial duration. In a further set of a forward/backward and a leftward/rightward articulation task, we additionally tested the influence of translational speed, including catch trials without articulation. We found a perceptual bias in translation direction in all three discrimination tasks. In the case of facing discrimination the bias was limited to short stimulus presentation. Our results suggest an interaction of articulation analysis with the processing of translational motion leading to best articulation discrimination when translational direction and speed match articulation. Moreover, we conclude that the global motion of the center-of-mass of the dot pattern is more relevant to processing of translation than the local motion of the dots. Our findings highlight that translation is a relevant cue that should be integrated in models of human motion detection.

Introduction

When we interact with other people and perceive their actions in space, we predict their future positions in the physical world in order to arrange an appropriate action ourself. Johansson (1973) showed that humans are remarkably good at recognizing human body motion even if kinematic information is reduced to a few light points attached to the head and the major joints of the body. This recognition is even possible if the point-light walker is presented for only 200 ms (Johansson, 1976). What distinguishes biological motion from object motion is the complex movement of the limb segments relative to each other. To isolate this kind of movement, referred to as articulation, biological motion has usually been investigated using point-light figures walking in place, as if on a treadmill. However, in the natural world, articulation results in a change of body location in space over time. This overall displacement, called translation, attaches the same speed and direction to all points of the body. Hence, real human body motion is a combination of articulation and translation, which completely describes its movement physically. For the point-light stimulus this means that the motion vector of each dot is composed of both an articulational and a translational component.

Translation is relevant to the perception of biological motion for two reasons. First, the physical principles of human body motion directly couple translation with articulation with respect to its direction (leftward/rightward or forward/backward) and speed. Thus, translation has a deterministic relationship to articulation. Second, the relevant information for the perceiver's action is not the movement of the body limbs itself, but the resulting and intended body position change in space over time. For these reasons, the translational cue is likely used in biological motion perception. If so, it would have to manifest itself in an influence of translation direction on the perception of articulation direction.

Processing of form and articulation

The question of how the articulation of a walking human is determined by the visual system has been the

focus of previous studies in biological motion perception. Whereas early findings emphasized local motion signals (Cutting, 1981; Johansson, 1973; Webb & Aggarwal, 1982), results of recent psychophysical studies (Beintema, Georg, & Lappe, 2006; Lu, 2010; McKay, Simmons, McAleer, & Pollick, 2009; Reid, Brooks, Blair, & van der Zwan, 2009; Theusner, de Lussanet, & Lappe, 2011; Thirkettle, Scott-Samuel, & Benton, 2010) strongly support the approach of a form-based analysis by first computing body posture and then body motion (Beintema & Lappe, 2002; Giese & Poggio, 2003; Lange, Georg, & Lappe, 2006; Theusner, de Lussanet, & Lappe, 2014). Appropriate posture-selectivity has been found in the occipital face area (Grossman & Blake, 2002; Michels, Lappe, & Vaina, 2005; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001), the fusiform body area (Michels et al., 2005; Peelen & Downing, 2005; Schwarzlose, Baker, & Kanwisher, 2005) and the extrastriate body area (Downing, Jiang, Shuman, & Kanwisher, 2001). In addition, neurons in the lower bank of the superior temporal sulcus (STS) and in the inferior temporal cortex (Singer & Sheinberg, 2010; Vangeneugden et al., 2011) respond to static postures of a body. Moreover, body motion analysis is assumed to be carried out in area STS (Bonda, Petrides, Ostry, & Evans, 1996; Grossman et al., 2000; Oram & Perrett, 1994, 1996; Thompson, Clarke, Stewart, & Puce, 2005; Vaina et al., 2001; Vangeneugden, Pollick, & Vogels, 2008), probably especially in its upper bank (Vangeneugden et al., 2011). Notably, area STS receives projections from both the ventral and the dorsal stream of the visual cortex, including the middle temporal area (MT; Ungerleider & Mishkin, 1982).

In a motion energy model of human body motion detection in 3D space, analogous to the standard model in luminance-based object motion detection (Adelson & Bergen, 1985; Burr, Ross, & Morrone, 1986; Reichardt, 1957; van Santen & Sperling, 1984; Watson & Ahumada, 1985), Theusner et al. (2014) described how the visual system may calculate biological motion and its features as they are used in discrimination tasks of many psychophysical experiments. According to their model, a presented body posture activates posture-selective neurons as stored 2D body template representations, each one to the extent of its matching with the presented posture. The most suitable template is then determined for each time step on the basis of the maximum responding posture-selective neuron. These maximum responses are summed up over the entire presentation duration to obtain facing direction. Subsequently, body motion energy is calculated by temporal integration of the sequence of activated template neurons. Equally, by maximum pooling of body motion energy for each time step and following summation over stimulus duration, articulation direc-

tion is derived. With respect to the temporal evolution of the calculation processes, Theusner et al. (2014) indicated that body motion-selective neurons discriminate articulation direction after a latency of about 200 ms, reaching a saturation level after about 400 to 600 ms after stimulus onset. In comparison, response latency to form is suspected to be around 119 ms (Oram & Perrett, 1996).

Processing of translation

As the translational component allocates correlated motion to the walking stimulus, it is likely to be analyzed by motion detectors in the dorsal pathway. Early stages of the motion stream contain local motion detectors corresponding to direction-selective neurons in the primary visual cortex (V1) and the middle temporal (MT) area. Area MT is involved in global motion detection as MT neurons of rhesus monkeys could extract motion on the basis of a minimum percentage (around 5%) of coherent motion among otherwise randomly moving noise dots (Newsome & Paré, 1988). Thus, motion direction of a simple dot pattern can be obtained by global motion processing on the basis of spatial pooling of local dot motion by linear summation. A subsequent processing stage in the motion pathway is the middle superior temporal area (MST), whose lateral-ventral region (MSTl) integrates intra- and extraretinal signals. Compared to neurons processing form and articulation of biological motion stimuli, MT and MST have rather short response latencies of around 70 ms (Schmolesky et al., 1998).

Combination of translation and biological motion

Research combining translation and biological motion is scarce. Neri, Morrone, and Burr (1998) compared the sensitivity of biological motion processing with that of simple translation and suggested that different integrative mechanisms exist for these two types of motion (but see Song & Perona, 2000). However, the walker in that study did not translate. Bidet-Ildei, Kitromilides, Orliaguet, Pavlova, and Gentaz (2014) examined biological motion with and without translation in newborn children. Comparing translating and nontranslating biological motion displays with translating and nontranslating scrambled displays, they revealed that the translational motion is the more salient aspect of the visual preference for biological motion in newborns. Thurman and Lu (2013) showed that added translational motion influenced animacy ratings of spatially scrambled biological motion stimuli. Furthermore, single cell responses in

the temporal cortex to biological motion stimuli including translation have been examined by Perrett, Harries, Chitty, and Mistlin (1990). They reported neurons being selective for one type of motion (articulation or translation). In a further study focusing on selectivity for body form from articulation and translation, Perrett, Harries, Benson, Chitty, and Mistlin (1990) found neurons discriminating body view using either translation or articulation and others using both of the two motion components tested separately. Therefore, they argued that articulation and translation have to be analyzed in two distinct channels that converge on a later stage. Furthermore, Oram and Perrett (1994) concluded that the integration of form, articulational and translational information must be the basis of sensitivity to human body motion.

Aim of this study

As a first step to find out how the translational component is integrated in biological motion processing the present study addresses the question of whether translation of a human walker influences the perception of its facing and its articulation direction, and, if so, how this influence is arranged in time. Translation direction could be a cue not just to articulation direction, but also to facing direction as facing and translation direction are correlated in our natural experience with the environment. We usually come more often upon people moving forward than moving backward so that translation could raise the probability for the corresponding facing direction. Hence, while articulation has a deterministic relationship to translation, facing has a probabilistic relationship to translation. If translation computation plays a role in perception of facing and or articulation direction, this influence could link to several possible neural mechanisms intervening on different processing levels of dynamic form analysis as described by Theusner et al. (2014).

Experiment 1

Experiment 1 contains three psychophysical experiments in which subjects respectively had to judge facing direction (leftward/rightward) and articulation direction (forward/backward and leftward/rightward) of a point-light walker in profile view with the translational component being varied. The two articulation tasks were included to test articulation perception on the basis of different required combinations of cues (absolute articulation direction in the leftward/rightward articulation task versus articulation direction

relative to facing direction in the forward/backward articulation task). In the case of an influence of translation on facing perception, translation consistent with facing direction should improve recognition compared to a stimulus walking in place, whereas translation inconsistent with facing direction should impair recognition in the facing discrimination task. Likewise, in the case of an influence of translation on articulation perception, translation consistent with articulation direction should have an advantage to a stimulus walking in place, which should have an advantage to translation inconsistent with articulation direction in the articulation discrimination tasks.

As translation allocates the same motion vector to each dot, it is likely to be processed faster than both form and articulation, so that a probable influence of translation on facing and articulation perception should be larger when facing and articulation signals are weak, i.e., in case of short stimulus presentation. Therefore, in all three experiments stimuli were presented for different durations to examine the temporal evolution of the probable influence of translation on facing and articulation perception. Following our assumptions, recognition should improve with higher stimulus duration (until the time when facing and articulation discrimination is saturated) compared with a shorter one (at the time when the threshold of facing and articulation is reached), whereas the impact of translation should become smaller, analogously.

In addition, we want to draw conclusions on how the translational component of biological motion is processed in the visual system. As translation adds correlated motion to the point-light walker, we hypothesize that translation is calculated on the basis of global motion signals, probably in area MT. Hence, we additionally manipulated the presence of local image motion on the basis of limited lifetime technique introduced by Beintema and Lappe (2002). If translation is obtained by local motion signals, we would expect an interaction of this factor with translation in terms of reduced influence of translation when local image motion is decreased.

Methods

Subjects

Two groups of 12 subjects participated. The first group performed all experiments with a stimulus duration of 200 ms. Subjects in this group were between 18 and 32 years old (22.8 years on average, all females). The second group of subjects was between 19 and 34 years of age (25.6 years on average, 10 females) and performed the experiments with stimuli presented for 800 ms. Some of the subjects were lab members who had ample experience with biological motion experi-

ments. The remaining subjects were students from the department. All subjects had normal or corrected-to-normal vision and were naive to the objective of the experiments.

Tasks

The three experiments partly differed in the composition of conditions but, with the respective features counterbalanced, resulted in the same overall stimulus sets. Each experiment included a two-alternative forced-choice task. In the facing task, we asked subjects to decide whether the walker was facing leftward or rightward. In the forward/backward articulation task, participants had to report whether articulation was forward or backward. The leftward/rightward articulation task also referred to articulation judgment, but in terms of discriminating between leftward and rightward articulation irrespective of facing direction.

Setup and stimuli

Stimuli consisted of computer-generated point-light walkers obtained by 3D motion-capture of a walking human. They respectively were displayed in middle screen position of a CRT monitor (40 × 30 cm, 1024 × 768 pixels, 120 Hz) in 2D orthographic profile view. Viewed from a distance of 60 cm in a dimly lit room, walkers measured 6.2° × 3.1° of visual angle (6.5 × 3.2 cm). They were composed of eight white points (0.3 cm in diameter, 0.3° of visual angle) on a dark gray background such that each of the eight limb segments (defined as upper and lower parts of the arms and legs) contained one dot at a random position within the respective limb segment. Frame duration was determined at 50 ms.

In all three tasks the walker could either face leftward or rightward with articulation directed either forward or backward (accordingly leftward or rightward) to a natural walking speed of 0.722 cycles/s.

Moreover, the point-light walker was translated horizontally either consistent or inconsistent with facing (facing task) or articulation (articulation tasks) or the figure was not translated at all (Figure 1). In the facing task, this means that a walker with translation defined as consistent was translated in the same direction as it was facing, regardless of whether its articulation (being balanced over trials) was directed forward or backward. In this experiment focus lay on the correspondence of facing and translation direction. Note that, as a consequence of articulation direction being counterbalanced, conditions of both consistent and inconsistent translation half contained walkers with articulation and translation defined in opposite directions. By doing so we ensured that only facing (and not articulation direction) could be the cue to the
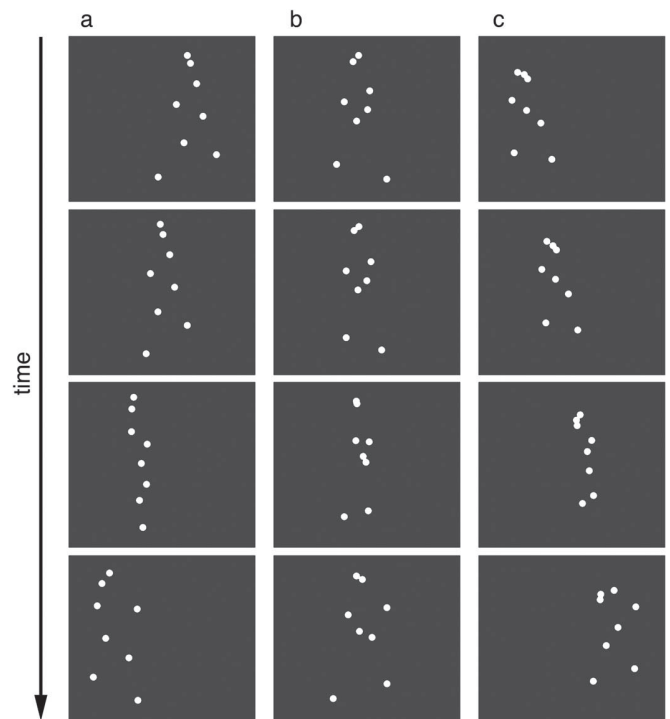

Figure 1. Manipulation of translation. Each column (a), (b), and (c) shows frame numbers 1, 6, 11, and 16 of a rightward facing point-light walker with backward articulation and infinite lifetime of points within the experiments of 800 ms stimulus duration. (a) Inconsistent translation for the facing task, consistent translation for the articulation tasks. (b) No translation. (c) Consistent translation for the facing task, inconsistent translation for the articulation tasks. Note that the walker always started from middle screen position and that the depicted dot positions are only an example as they were randomly selected with every single trial.

direction discrimination in demand. For consistent translation in the articulation tasks, analogously, translation direction was equal to articulation direction, thus, both directed either forward or backward. Accordingly, this means that in the articulation tasks, conditions of consistent translation contained only walkers with natural articulation-translation coupling regardless of absolute facing direction that was balanced over trials. By contrast, in conditions of inconsistent translation, only walkers were presented with the translation and articulation aligned in opposite directions. At the beginning of the trial, the point-light walker was depicted in the middle of the screen and then, starting from there, changed its position from frame to frame either to the left or to the right side of the monitor. Translational speed matched the natural physical coupling to articulation. In conditions of no translation the translational component was subtracted in all three experiments giving the impression of a
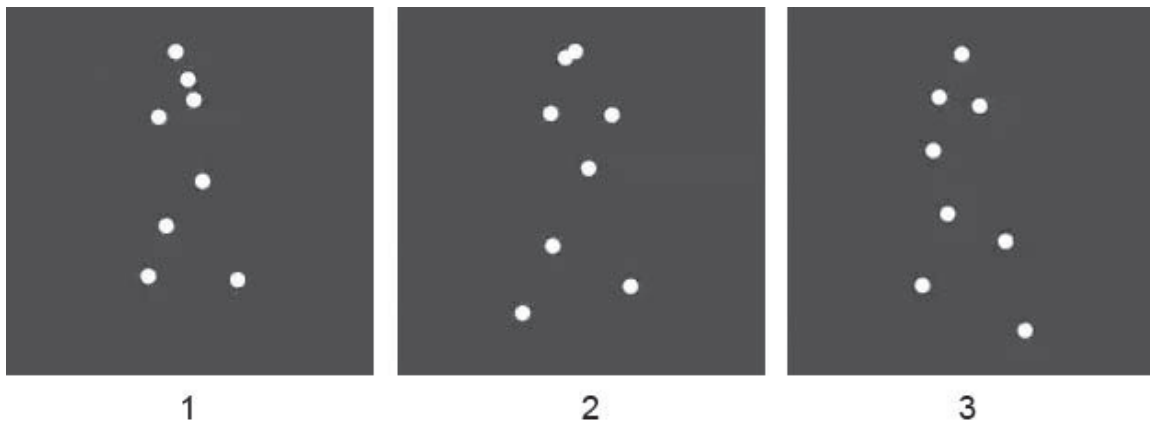
Movies 1–3. Movies of example stimuli with rightward facing, forward articulation and point lifetime of 1 frame. (1) Consistent translation. (2) No translation. (3) Inconsistent translation for the facing and the articulation tasks. Click on the image to view the movie. In the experiments the walker was shown for 200 ms or 800 ms.

person walking on a treadmill. Example stimuli can be seen in Movies 1 to 3.

In order to manipulate the amount of local image motion we additionally varied lifetime of points (Figure 2). This defines the number of frames after which the light-points were removed and replaced each at a new random location within the respective limb segment. Just as the initial positions of the points these following positions were randomly selected in each single trial presentation. Limiting lifetime disturbs a point's correspondence in body position between different frames and, thus, reduces the continuous motion signal that each single dot carries in terms of the apparent motion vector that becomes established over successive frames. Consequently, the amount of local motion information is reduced while global form and motion of the walker are left intact. Lifetime of points could be one frame (no local motion), two frames (the minimum to compute local motion), or infinite (full local motion). For a lifetime of two frames points were relocated asynchronously over successive frames to achieve a temporally constant amount of local image motion.

The two groups of subjects participated in the experiments for different durations of stimulus presentation. One of them was determined at the threshold for articulation discrimination (200 ms, according four frames) and the other one at a duration to which the saturation level for articulation discrimination is reached (800 ms, according 16 frames). Even though facing discrimination is overall easier than articulation discrimination (Beintema et al., 2006; Lange & Lappe, 2006; Wittinghofer, de Lussanet, & Lappe, 2012), we chose the same stimulus durations for the facing task for the purpose of comparing the results among the three experiments afterwards. With 200 ms stimulus duration, the walker performed only 14% of a gait cycle (defined about two walking steps) while, in the case of translation, being displaced 1.0° of visual angle (1.0 cm)

to the left or to the right. With stimulus duration of 800 ms, it performed roughly a half gait cycle with displacement of 3.8° of visual angle (4.0 cm) in the case of translation. Starting phase of the walking cycle in terms of the first depicted posture was balanced over trials by four different start phases. Each experimental session comprised three experiments each composed of 144 trials (2 facing directions [leftward/rightward] $\times$ 2 articulation directions [forward/backward or leftward/ rightward] $\times$ 3 translation directions [consistent/none/ inconsistent] $\times$ 3 lifetimes of points [1/2/infinite] $\times$ 4
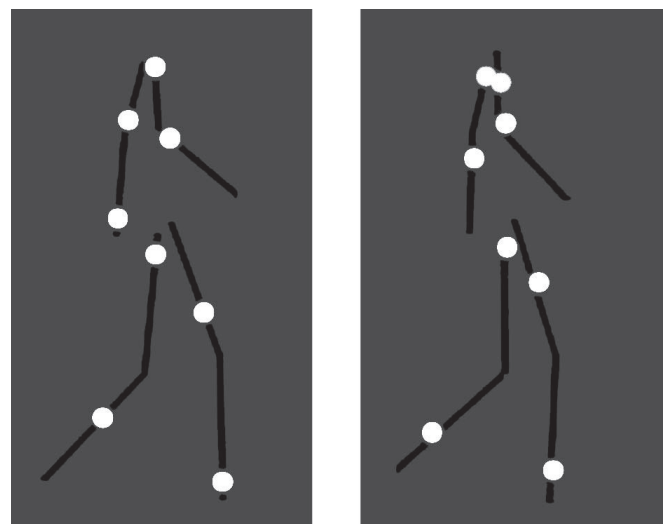

Figure 2. Manipulation of local image motion. With lifetime of points being reduced to one frame all dots are relocated to another randomly selected position within each limb segment so that trajectory information is disturbed. The black stick figure is only shown for demonstration and was not part of the stimulus in the experiments.

starting phases; 2 stimulus durations [200 ms/800 ms] as between-subject factor).

Procedure

Before data collection we presented a demo version of classic walkers (with 12 point-lights attached to the body's major joints) to familiarize participants with the difference between articulation (introduced as "movement of the limbs relative to each other") and translation (introduced as "overall movement across the screen") as well as the difference between the questions of facing direction, forward/backward articulation and leftward/rightward articulation in general. Each experiment was preceded by a block of 36 practice trials without feedback to stabilize performance. On the basis of the introduced descriptions participants were explicitly instructed to judge the respective feature regardless of other features (for example, to judge leftward/rightward articulation regardless of facing and translation) as exactly as possible. They also were informed that reaction time does not play a role and, in case of uncertainty, asked to choose the alternative they rather tended to. Response was indicated after disappearance of the stimulus either by pressing the "left" or "right" (in the facing and leftward/rightward articulation task) or the "up" or "down" (in the forward/backward articulation task) arrow key of the keyboard without any time constraint. Upon button response, the next trial started after a pause of 500 ms. The order of experiments was balanced between the participants with the respective trials randomly interleaved. In sum, the entire data collection session took about 30 min for 200 ms stimulus duration and about 50 min for 800 ms stimulus duration.

Data analysis

Data analysis was based on d' values, which, in comparison to simple mean correct rates, quantify discriminability adjusted for response bias. The problem of proportions of 0 and 1 leading to infinite d' values was solved by replacing them by $\frac{1}{2n}$ and $1 - \frac{1}{2n}$ respectively, where n is the number of trials included (Macmillan & Creelman, 2005). By doing so, perfect and null accuracy were treated as if participants made a half hit or error. For each experiment, mixed analysis of variance was calculated on d' with the three factors translation, lifetime of points, and stimulus duration (3 $\times$ 3 $\times$ 2 ANOVA with stimulus duration as a between-subject factor). If appropriate, follow-up ANOVAs were computed and, subsequently, paired and unpaired t tests were applied for post-hoc pairwise comparisons corrected by the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). Additionally, results were tested against chance level using one-

sample t tests, also used with the Benjamini-Hochberg correction.

Results

Figure 3 presents the participants' performance for all three discrimination tasks.

Facing task

In comparison across all experiments, participants reached best performance in the facing discrimination task. Discriminability was generally better with 800 ms than with 200 ms stimulus duration, main effect of stimulus duration, $F(1, 22) = 45.27$, $p < 0.001$, and decreased when lifetime of points was prolonged, main effect of lifetime of points, $F(2, 44) = 15.38$, $p < 0.001$. Also translation had an influence on performance, main effect of translation, $F(2, 44) = 7.02$, $p = 0.002$, that was manifested differently for the two stimulus durations, interaction of stimulus duration with translation, $F(2, 44) = 5.06$, $p = 0.010$, but equally for the different lifetimes of points, no interaction of lifetime of points with translation, $F(4, 88) = 0.41$, $p = 0.801$. Stimulus duration determined the effect of translation on performance such that, with 200 ms stimulus duration, participants showed a tendency to perform better when translation was consistent with facing direction than for walking-in-place stimuli (but no significant difference in post-hoc t tests aggregated over lifetime of points, $p = 0.602$). However, performance with inconsistent translation was impaired against both consistent and no translation ($p \leq 0.010$), even dropping to chance level for two frames and infinite lifetime of points ($p \geq 0.083$; all other conditions $p \leq 0.005$; Figure 3). Contrary to this, for 800 ms stimulus duration, translation rather disturbed facing discrimination irrespectively of whether it commenced in the same or opposite direction as facing (significant difference between no translation and each consistent and inconsistent translation, $p \leq 0.021$; no significant difference between consistent and inconsistent translation, $p = 0.927$). Except for the two mentioned all pairwise comparisons revealed significant differences ($p \leq 0.021$).

Forward/backward articulation task

The articulation discrimination tasks were generally more difficult for participants than the judgment of facing direction. However, response patterns in the forward/backward task also showed a higher discrimination rate with 800 ms than with 200 ms stimulus duration, main effect of stimulus duration, $F(1, 22) = 19.64$, $p < 0.001$, and an influence of translation, main effect of translation, $F(2, 44) = 28.48$, $p < 0.001$. The
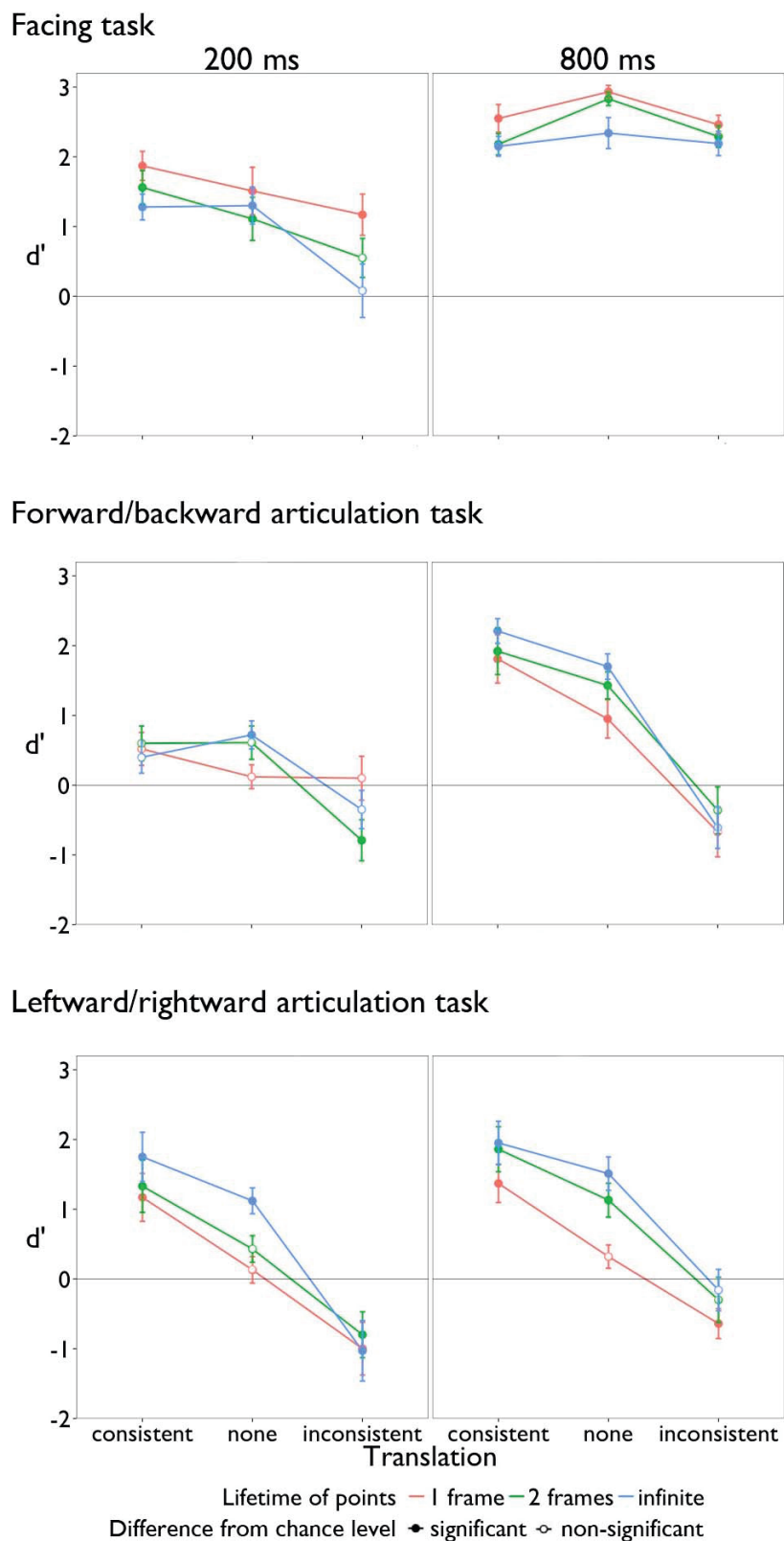
Figure 3. Experiment 1: Means and standard errors in *d'* values. *d' = 0* indicates chance level performance, *d' ≥ 3* almost saturated discriminability.

latter was affected by stimulus duration, interaction of stimulus duration with translation, $F(2, 44) = 6.34$, $p = 0.015$, such that, with 200 ms stimulus duration, performance of consistent translation was similar to that for walking-in-place stimuli whereas inconsistent translation led to a small decline in performance. In several conditions of 200 ms stimulus duration, performance lay around chance level ($p \geq 0.056$; all other conditions $p \leq 0.047$; Figure 3). By contrast, within 800 ms stimulus duration, one could observe a benefit from consistent translation in comparison to no translation and a comparably strong decline in performance for inconsistent translation that dropped to chance level for all three lifetimes of points ($p \geq 0.094$). Additionally, the translational influence was affected by lifetime of points, interaction of lifetime of points with translation, $F(4, 88) = 3.53$, $p = 0.010$, apparently underlying deviating response pattern within lifetime of points of one frame and 200 ms stimulus duration, but not in terms of reduced influence of translation with decreasing lifetime of points in general.

Leftward/rightward articulation task

The leftward/rightward articulation task revealed a similar performance range and response pattern as the forward/backward articulation task with stimulus duration of 800 ms. Subjects could better recognize absolute articulation direction when stimuli were presented for 800 ms than when presented for 200 ms, main effect of stimulus duration, $F(1, 22) = 6.83$, $p = 0.016$. In contrast to the facing task, an increase in lifetime of points had a positive effect on discriminability, main effect of lifetime of points, $F(2, 44) = 19.87$, $p < 0.001$. Moreover, it became apparent that translation is relevant for articulation perception, main effect of translation, $F(2, 44) = 32.12$, $p < 0.001$, irrespective of stimulus duration, no interaction of stimulus duration with translation, $F(2, 44) = 0.11$, $p = 0.773$. A weak interaction of translation with lifetime of points was found, $F(4, 88) = 3.00$, $p = 0.049$, but was not confirmed by the two follow-up ANOVAs, $F(4, 44) = 2.28$, $p = 0.075$ for 200 ms stimulus duration, $F(4, 44) = 1.31$, $p = 0.280$ for 800 ms stimulus duration. Hence, for 200 ms as well as for 800 ms stimulus duration, participants reached higher discrimination rate in comparison to treadmill walking stimuli when translation was consistent to articulation direction and, on the other hand, reached lower discrimination rate when translation was inconsistent to articulation direction (post-hoc t tests aggregated over lifetime of points; $p \leq 0.041$ within stimulus duration of 200 ms; $p \leq 0.002$ within stimulus duration of 800 ms). Even though, it emerged that subjects had difficulties in judging articulation direction even for stimuli walking in place

as performance in some conditions within missing translation (and also within inconsistent translation for 800 ms stimulus duration) was at chance level ($p \geq 0.059$; all other conditions $p < 0.050$; Figure 3). For inconsistent translation, participants judged articulation direction systematically beneath chance level for stimulus duration of 200 ms in general and for 800 ms with lifetime of points of one frame ($p < 0.050$).

Discussion

Influence of translation and its temporal evolution

In all three experiments we found better discrimination for longer stimulus duration and an influence of the translational component on perception. The latter, however, was different in each task, suggesting different influences on the form and motion levels of biological motion perception.

Level of form analysis. Evidence for the role of translation on the level of global form analysis was provided by the facing task. With the presentation duration of 200 ms, results were significantly different from chance level even without translation but below saturated performance (72.0% on average). Presentation over 800 ms of stimuli articulating in place showed almost saturated discriminability (94.4% on average). This is in line with our assumptions as well as with findings in the single cell study of Oram and Perrett (1996) about neurons in the macaque monkey's STS responding with a latency of 119 ms to human body postures.

Our data confirm the hypothesis that facing perception is biased in the translational direction, but this influence is restricted to the case of short stimulus presentation, i.e., when form information is weak. As a consequence, translation must be determined with a latency of less than 200 ms which is consistent with response latencies in MT and MST (Schmolesky et al., 1998). Concerning the temporal evolution of the translational influence on perception of form, though, results do not indicate a decrease with longer stimulus duration but rather another characteristic of the translational influence. When the form signal is strong (owing to longer presentation duration), translation just impairs facing judgment irrespective of whether it is directed in the same or opposite direction of facing. If perception of facing direction is based predominantly on form analysis, a translation of the figure over time would simply be a further complication of the task. The results at 200 ms, however, show a dependence on translation. This may be explained, in case of limited form information, if perception is based on a probabilistic connection between facing and translation (people usually face in the direction in which they move) such that inconsistent translation introduces response errors. However, this may be an effect on the

level of response selection rather than on the level of biological motion analysis.

Level of articulation analysis. The question of whether calculation of translation plays a role in articulation analysis was addressed by two articulation tasks. Our findings of increasing discriminability with longer stimulus duration in both experiments agree with our assumptions and the model of Theusner et al. (2014) concerning the temporal evolution of articulation discrimination on neural level. Moreover, both experiments reveal an impact of the translational component on the perception of articulation direction.

In the first place, for the forward/backward task with 200 ms stimulus duration, perception did not benefit from consistent translation in comparison to missing translation but suffered from inconsistent translation. Nevertheless, as performance within stimulus duration of 200 ms lay around chance level in most conditions, we are cautious in interpreting this interaction of stimulus duration and translation concerning the temporal evolution of the translational influence. It rather seems that, with stimulus duration being limited to 200 ms, perception of articulation was just too difficult. In contrast, at 800 ms stimulus duration perception of articulation was possible in the no translation condition, and consistent translation further improved perception whereas inconsistent translation impaired perception. This response pattern cannot be explained by participants simply indicating forward articulation in case of natural walkers (consistent translation) and backward articulation in case of unnatural walkers (inconsistent translation) because, if that were the case, performance should lie around chance level in both translational conditions. Therefore, the discriminability within 800 ms stimulus duration clearly shows that the translational component is an interacting factor in perception of articulation direction, which biases perception in translational direction.

The leftward/rightward articulation task also showed an influence of translation on articulation perception in terms of higher discriminability with consistent translation and lower discriminability with inconsistent translation in comparison to no translation. Moreover, in the leftward/rightward articulation task the influence of translation was independent of stimulus duration such that translation affected articulation perception similarly within 200 ms and 800 ms stimulus duration.

In order to explain the differences between the two articulation tasks at 200 ms stimulus duration one has to remember that they require different combinations of cues. The leftward/rightward articulation task calls for an absolute articulation direction in space, irrespective of whether the point-light walker faces leftward and walks forward or faces rightward and walks backward, for example. Both cases would request the same decision (leftward). Translation, in the consistent condition, directly matches the requested decision in each case. In the inconsistent condition, in contrast, translation is always opposite to the correct decision. Therefore, if participants based their response merely on the translation direction, good performance would have been expected for consistent translation, chance level performance for no translation and systematic errors for inconsistent translation. In general, the pattern of data in the leftward/rightward articulation task cannot exclude this possibility.

In the forward/backward task, in contrast, translation on its own is not informative for the response. In the consistent condition of forward articulation, for example, walkers facing left translate to the left while walkers facing right translate to the right. Both cases request the decision "forward", irrespective of translation direction. Therefore, if articulation perception is difficult, chance level performance would be expected in any case. Overall, the data at 200 ms is consistent with this, since performance is mostly at chance. However, there was a slight but significant influence of translation. Such an influence could be explained if participants used translation direction in comparison with facing direction for their response, assuming that articulation was forward when facing and translation were codirectional and backward otherwise. This would point towards an interaction at the response level. However, it could make only a minor contribution to the data at 200 ms since the influence of translation was small.

Local versus global image motion

The effect of translation in the facing and the leftward/rightward articulation experiment was independent of lifetime of the dots. The interaction of translation with lifetime of points in the forward/backward articulation task did not result in a smaller effect of translation with reduced lifetime of points. Overall, therefore, translation had an effect on biological motion perception in all lifetime conditions, even at a lifetime of only one frame. We conclude, therefore, that the global translation of the center-of-mass of the dot pattern is the basis of translation processing. If translation is computed separately to articulation, this could take place in area MT of the dorsal pathway that is connected to area STS (Ungerleider & Mishkin, 1982). However, one has to consider a limitation of our local motion manipulation within the translational conditions. Without translation, a lifetime of one frame efficiently disturbs the spatial correspondence between the dots in successive frames (Beintema et al., 2006). With translation, however, the additional horizontal motion of the entire

body leads to increased consistency of successive dot positions with respect to the horizontal axis so that the correspondence between dots across frames might be higher in translational conditions. Nonetheless, since we found a main effect of lifetime of points in both the facing and the leftward/rightward articulation experiment, the manipulation must have been successful.

Shorter point lifetime led to better performance in the facing task. This may be explained by an increase in form information due to a more rapid sampling of the limb positions in this case. In contrast, the increase in performance with lifetime in the leftward/rightward articulation task suggests that local motion is used as an additional cue in this task.

Experiment 2

In Experiment 2 we collected further evidence that the translational influence on articulation direction that was observed in Experiment 1 is due to an interaction on the neural level of processing of human movement rather than an interaction on the response level. We presented a forward/backward articulation and a leftward/rightward articulation task equivalent to those in Experiment 1 but manipulated speed of translation relative to articulation. If the translational influence arises from an interaction between translation and articulation on the neural level, we would expect best discrimination for translation speed being consistent to articulation speed, and poorer discrimination when translational speed is higher or lower than articulational speed. In contrast, if the translational influence is due to response level interaction, performance should be unaffected by translational speed or even improve with increasing translational speed as long as it is in the consistent direction.

In addition, we included catch trials in which only translation but no articulation was shown. If the interaction between translation and articulation occurs at response level, responses in these catch trials should be determined by translation direction. If, in contrast, translation and articulation interact on a neural level, responses in the articulation task should be at chance level irrespective of translation.

Methods

Subjects

The 13 participants were between 18 and 51 years of age (28.2 years on average, 10 females). Some of them were lab members. All subjects had normal or corrected-to-normal vision and, apart from the authors JM and ML, were naive to the objective of the experiments.

Tasks

The forward/backward and the leftward/rightward articulation task were the same as in Experiment 1.

Setup and stimuli

Point-light walkers were the same as in Experiment 1 and could either face leftward or rightward and articulate forward or backward (accordingly leftward or rightward). Lifetime of points was fixed at one frame as, following Experiment 1, reduced local motion does not impair the effect of translation. Moreover, since Experiment 1 revealed that discrimination of forward/backward articulation was too difficult at 200 ms, all stimuli were presented for 800 ms.

The point-light figure translated horizontally either consistent or inconsistent with articulation direction regardless of facing direction. This ensured that only articulation (and not translation direction) was a reliable cue to solve the articulation discrimination task. Speed of translational motion was varied as a factor of articulational speed and, thus, could be either 0 (no translation), 0.5 (half of articulational speed with 2.4°/s), 1 (same as articulational speed, matching the natural physical coupling to articulation with 4.8°/s), or 2 (double of articulational speed with 9.5°/s). Hence, with translational speed of 0, the point-light walker was walking in place, whereas it was displaced horizontally 1.9, 3.8, or 7.6° of visual angle (2, 4, or 8 cm) to the left or to the right with translational speed of 0.5, 1, or 2. In catch trials articulation was removed as cue. In these trials a static body posture of a point-light figure was translated across the screen.

In sum, each experimental session comprised two experiments with 128 trials each (2 facing directions [leftward/rightward] × 2 articulation directions [forward/backward or leftward/rightward] × 2 translation directions [consistent/inconsistent] × 4 translation speeds [0/0.5/1/2] × 2 trial types [standard/catch] × 2 repetitions).

Procedure

Participants received the same instruction as in Experiment 1. The order of trials was randomized with an intertrial interval of 500 ms. The experiment lasted about 25 min for each subject. The order of experiments was counterbalanced.

Data analysis

For each experiment, we applied a one-way analysis of variance on d' for the standard and for the catch trials, respectively. Translation was used as a general factor with the levels −2, −1, −0.5 (translational speeds for inconsistent translation), 0 (no translation), 0.5, 1,

and 2 (translational speeds for consistent translation) for the standard trials. For the catch trials (without articulation) levels were restricted to 0.5, 1, and 2 (translational speeds without translation direction) and *d'* values were calculated based on conformity with translation direction (forward/backward translation in the forward/backward articulation task and leftward/rightward translation in the leftward/rightward articulation task). Moreover, paired *t* tests for pairwise comparisons and one-sample *t* tests against chance level were computed, both corrected by the Benjamini-Hochberg method.

Results

Forward/backward articulation task

In the standard trials of the forward/backward articulation task, discriminability depended highly on translation, main effect of translation, $F(6, 72) = 8.22$, $p = 0.005$. Figure 4 presents the data. Responses were at chance level performance for all conditions of inconsistent translation ($p \geq 0.348$). Performance was above chance level for translational speed 0 (treadmill-walking figures) and all conditions of consistent translation ($p \leq 0.048$). At double translational speed in consistent direction performance was significant lower than for natural speed matched to articulation ($p = 0.009$).
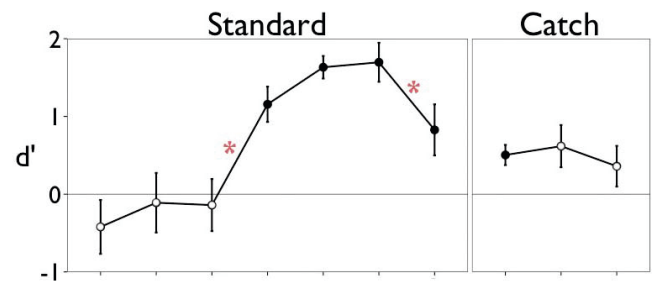
In the catch trials, the frequency of responses conforming to translation direction was not affected by translational speed, no significant main effect of translation, $F(2, 24) = 0.60$, $p = 0.559$; $p = 0.630$ for all three pairwise comparisons. Beyond that, subjects' judgement of articulation direction was determined by translation direction in about 50% of cases for natural and double translational speed ($p \geq 0.063$ for conditions of 1 and 2) and slightly above 50% for half translational speed ($p = 0.007$ for condition of 0.5).

Leftward/rightward articulation task

The response pattern in the leftward/rightward articulation task was similar to that in the forward/backward articulation task (Figure 4). Performance was influenced by translation, main effect of translation, $F(6, 72) = 5.11$, $p = 0.013$. Discrimination was at chance level for all conditions of inconsistent translation ($p \geq 0.086$) and above chance level for conditions of no translation and half and natural translational speed in consistent direction ($p \leq 0.003$). Performance dropped to chance level for double translational speed in consistent direction ($p = 0.695$; significant difference between conditions of 1 and 2, $p = 0.035$).

Like in the forward/backward task, the frequency of responses conforming to translation direction in the
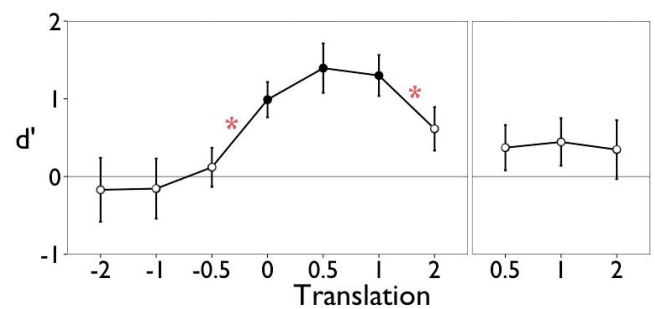
Figure 4. Experiment 2: Means and standard errors in *d'* values. *d'* = 0 indicates chance level performance, *d'* ≥ 3 almost saturated discriminability. Asterisks indicate significant differences between conditions.

catch trials was independent of translational speed, no significant main effect of translation, $F(2, 24) = 0.05$, $p = 0.888$; $p = 0.930$ for all three pairwise comparisons. The rate of responses conforming to translation was about 50% for all translational speeds ($p \geq 0.341$).

Discussion

The results of Experiment 2 are in line with the results of the articulation tasks for 800 ms of Experiment 1. In both the forward/backward and the leftward/rightward articulation task, inconsistent translation impaired perception of articulation direction such that performance dropped to chance level. Experiment 2 additionally revealed that discriminability within inconsistent translation is independent of the relation between articulational and translational speed. Confirming Experiment 1, response patterns indicated a higher discrimination rate for treadmill-walking figures in comparison with inconsistent translation.

For consistent translation, both articulation tasks showed that perception of articulation relies on translational speed. Performance was best for a translational speed that matched the physical articulation-translation coupling or slightly slower. Performance was impaired

when translation was twice as fast as articulation. This is in line with the expectation for a perceptual translational influence on the level of neural interaction of the two motion components. This conclusion is also supported by the catch trial data in both the forward/backward and leftward/rightward articulation task revealing that, in case of absence of actual articulation, responses did not conform to translation direction.

General discussion

The broader purpose of the present study was to extend the question of how the articulational aspect of biological motion is processed in the visual system to more real human movement as it occurs in the natural world, therefore including translation. To determine how translation calculation could be integrated in biological motion analysis we investigated the influence of translation on facing and articulation perception on the basis of different perceptual discrimination tasks. Additionally, we examined whether this influence is moderated by time and the amount of local image motion (Experiment 1) and whether it occurs on the level of neural processing of human movement (Experiment 2).

In Experiment 1 presented with 200 ms stimulus duration, we found an effect of translation direction on facing discrimination in terms of better performance when translation was consistent to facing direction in comparison to a treadmill-walking figure. In contrast, translation that was inconsistent to facing direction led to poorer performance. With stimulus duration of 800 ms, discriminability was reduced with both consistent and inconsistent translation against the treadmill-walking stimulus. The influence of translation direction occurred irrespective of lifetime of points with both trial durations. In the articulation discrimination tasks, responses were also biased in translational direction except for the forward/backward articulation task with 200 ms stimulus duration where performance was around chance level. Reduction of lifetime of points did not impair the translational influence on articulation discrimination. Additionally, Experiment 2 revealed that articulation discrimination is most successful when consistently directed translational speed matches articulation or is slightly slower. In catch trials in which articulation was subtracted, responses were independent of translation direction and speed.

The results of Experiments 1 and 2 clearly indicate that the process of human motion processing takes into account the change of body localization in space over time that is physically coupled to the articulational movement. From Experiment 1 we conclude that the translational component of biological motion is processed with a latency of less than 200 ms on the basis of global motion detection, thus, by the center-of-mass of the dot pattern in the case of a point-light figure.

Moreover, translation is a contributory factor in perception of facing and articulation by driving perception in the translational direction. Experiment 1 revealed that on the level of form analysis, this translational influence on perception is restricted to short stimulus presentation. At longer stimulus duration translation rather disturbed facing perception. Consequently, these data rather point to a separate analysis of form and translation rather than an interaction at the neural level. Nevertheless, these two signals could come into competitive interaction at the response level such that, for a lack of form information due to short stimulus duration, perception is dominated by the translation percept. This would be an economic mechanism with respect to the probabilistic relationship of facing and translation direction in natural experience.

In contrast, the translational effect in the articulation tasks of Experiment 1 was maintained up to stimulus duration of 800 ms. Furthermore, Experiment 2 revealed that this effect cannot be explained on the level of response. Thus, a neural convergence of the translational and the articulational processing stream in visual analysis seems likely. This could be realized by the projections of the dorsal pathway into area STS (Ungerleider & Mishkin, 1982) and would be in line with the conclusions of Perrett, Harries, Benson, et al. (1990) about a combination of the two motion signals on neural level. However, apart from an initially separate processing of the two motion components, results of the articulation tasks could also be due to an integrative processing of articulation and translation in terms of a combined selectivity for both, for example by body motion detectors being selective for the natural combination of articulation and translation. This would also be consistent with the findings of Experiment 2 illustrating that a natural walker with articulation and translation of consistent direction and speed is optimal for discriminating articulation direction. As the question of the underlying neural mechanisms has implications for recent models of biological motion perception, it should be the focus of future investigations.

Keywords: human walking, biological motion processing, translation

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*(2), 284–298.

Beintema, J. A., Georg, K., & Lappe, M. (2006). Perception of biological motion from limited lifetime stimuli. *Perception & Psychophysics, 68*(4), 613–624.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences, USA, 99*(8), 5661–5663.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57*(1), 289–300.

Bidet-Ildei, C., Kitromilides, E., Orliaguet, J.-P., Pavlova, M., & Gentaz, E. (2014). Preference for point-light human biological motion in newborns: Contribution of translational displacement. *Developmental Psychology, 50*(1), 113–120.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16*(11), 3737–3744.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London B, 227,* 249–265.

Cutting, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*(1), 71–87.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470–2473.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179–192.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*(6), 1167–1175.

Grossman, E. D., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12*(5), 711–720.

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

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research, 38,* 379–393.

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, 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*(11), 2894–2906.

Lu, H. (2010). Structural processing in biological motion perception. *Journal of Vision, 10*(12):13, 1–13, doi:10.1167/10.12.13. [PubMed] [Article]

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

McKay, L. S., Simmons, D. R., McAleer, P., & Pollick, F. E. (2009). Contribution of configural information in a direction discrimination task: Evidence using a novel masking paradigm. *Vision Research, 49*(20), 2503–2508.

Michels, L., Lappe, M., & Vaina, L. M. (2005). Visual areas involved in the perception of human movement from dynamic form analysis. *NeuroReport, 16*(10), 1037–1041.

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature, 395,* 894–896.

Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area MT. *Journal of Neuroscience, 8*(6), 2201–2211.

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience, 6*(2), 99–116.

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.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93*(1), 603–608.

Perrett, D. I., Harries, M., Chitty, A. J., and Mistlin, A. J. (1990). Three stages in the classification of body movements by visual neurones. In H. B. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding* (pp. 94–108). Cambridge, UK: Cambridge University Press.

Perrett, D. I., Harries, M. H., Benson, P. J., Chitty, A. J., and Mistlin, A. J. (1990). Retrieval of structure from rigid and biological motion: An analysis of the visual response of neurons in the macaque

temporal cortex. In T. Troscianko & A. Blake (Eds.), *AI and the eye* (p. 181–201). Chichester, UK: Wiley.

Reichardt, W. (1957). Autokorrelations-Auswertung als Funktionsprinzip des Zentralnervensystems [Translation: Autocorrelation analysis as functional principle of the central nervous system]. *Zeitschrift für Naturforschung, 12b*(7), 448–457.

Reid, R., Brooks, A., Blair, D., & van der Zwan, R. (2009). Snap! Recognising implicit actions in static point-light displays. *Perception, 38*(4), 613–616.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79,* 3272–3278.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25*(47), 11055–11059.

Singer, J. M., & Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *Journal of Neuroscience, 30*(8), 3133–3145.

Song, Y., & Perona, P. (2000). A computational model for motion detection and direction discrimination in humans. *Proceedings, Workshop on Human Motion, 2000,* 11–16.

Theusner, S., de Lussanet, M., & Lappe, M. (2011). Adaptation to biological motion leads to a motion and a form aftereffect. *Attention Perception & Psychophysics, 73*(6), 1843–1855.

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. *Journal of Neuroscience, 34*(3), 909–921.

Thirkettle, M., Scott-Samuel, N. E., & Benton, C. P. (2010). Form overshadows 'opponent motion' information in processing of biological motion from point light walker stimuli. *Vision Research, 50*(1), 118–126.

Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *Journal of Neuroscience, 25*(39), 9059–9066.

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

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. Mansfield (Eds.), *The analysis of visual behavior* (p. 549–586). Cambridge, MA: MIT Press.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences,* USA, 98(20), 11656–11661.

Vangeneugden, J., De Maziere, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *Journal of Neuroscience, 31*(2), 385–401.

Vangeneugden, J., Pollick, F., & Vogels, R. (2008). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cerebral Cortex, 19*(3), 593–611.

van Santen, J. P., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*(5), 451–473.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual motion sensing. *Journal of the Optical Society of America A, 2*(2), 322–341.

Webb, J. A., & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Computer Vision and Image Understanding, 19,* 107–130.

Wittinghofer, K., de Lussanet, M., & Lappe, M. (2012). Local-to-global form interference in biological motion perception. *Attention, Perception & Psychophysics, 74*(4), 730–738.