Deriving motor primitives through action segmentation

Paul E. Hemeren* and Serge Thill

School of Humanities and Informatics, University of Skövde, Skövde, Sweden

The purpose of the present experiment is to further understand the effect of levels of processing (top-down vs. bottom-up) on the perception of movement kinematics and primitives for grasping actions in order to gain insight into possible primitives used by the mirror system. In the present study, we investigated the potential of identifying such primitives using an action segmentation task. Specifically, we investigated whether or not segmentation was driven primarily by the kinematics of the action, as opposed to high-level top-down information about the action and the object used in the action. Participants in the experiment were shown 12 point-light movies of object-centered hand/arm actions that were either presented in their canonical orientation together with the object in question (top-down condition) or upside down (inverted) without information about the object (bottom-up condition). The results show that (1) despite impaired high-level action recognition for the inverted actions participants were able to reliably segment the actions according to lower-level kinematic variables, (2) segmentation behavior in both groups was significantly related to the kinematic variables of change in direction, velocity, and acceleration of the wrist (thumb and finger tips) for most of the included actions. This indicates that top-down activation of an action representation leads to similar segmentation behavior for hand/arm actions compared to bottom-up, or local, visual processing when performing a fairly unconstrained segmentation task. Motor primitives as parts of more complex actions may therefore be reliably derived through visual segmentation based on movement kinematics.

Keywords: action representation, event segmentation, motor primitives, mirror neurons, point-light displays, biological motion, action recognition, motor cognition

INTRODUCTION

Human actions have a spatiotemporal structure that can be accessed when we execute our own actions and when we view the actions of others. Understanding the structure of human actions allows us to identify and predict, or to see, the intentions of others (e.g., Blakemore and Decety, 2001; Saxe et al., 2004; Iacoboni et al., 2005). This ability to see a pattern of human motion as a coherent whole and not just as a complex pattern of movement of the arms and legs is referred to as epistemic visual perception (Jeannerod and Jacob, 2005). The meaning of actions is tied to the conceptual knowledge associated with a given pattern of bodily motion. Conceptual knowledge in turn includes knowledge about the goals of the movement as well as its sensory–motor patterns in motor execution and the visual recognition of actions. This linkage between perceptual and conceptual knowledge is apparent in tasks that demonstrate interference effects between perceptual and conceptual tasks (van Dantzig et al., 2008). The research presented here further addresses the relationship between perceptual and conceptual knowledge in the context of event segmentation (Zacks and Swallow, 2007).

ACTION REPRESENTATION, MOTOR PRIMITIVES, AND EVENT STRUCTURE

Motor execution and the visual recognition of actions appear to have a common neurological basis in mirror neurons in the primate brain, which become activated when an individual performs certain actions *and* when an individual observes the actions of another person performing the actions (e.g., di Pellegrino et al., 1992; Rizzolatti et al., 1996, 2006; Buccino et al., 2004). Mirror neurons have been described as a critical neurophysiological basis for internal models for action representation and for mediating the coupling between perception and action (Pozzo et al., 2006). These internal models have in other work been referred to as action prototypes (e.g., Pollick et al., 2001; Giese and Lappe, 2002; Pollick, 2004; Hemeren, 2008), motor schemata (Grafton et al., 1997), and motor prototypes (Borghi and Riggio, 2009). The ability to imagine, or simulate, movement relies on activation of cortical areas (e.g., primary motor cortex, the supplementary motor area, and the premotor cortex) that are involved in the execution of actions (Jeannerod, 1995; Michelon et al., 2006). One central idea regarding the structure and organization of the internal models concerns the existence of motor primitives that can be flexibly combined to create complex action sequences (e.g., Mussa-Ivaldi and Bizzi, 2000; Thoroughman and Shadmehr, 2000; Poggio and Bizzi, 2004; Chersi et al., 2006). The combination of motor primitives can also create hierarchical representations that allow generalization over specific situations (Poggio and Bizzi, 2004). Conversely, action hierarchies can also be understood on the basis of their component structure.

In order to understand our ability to represent actions, we need to understand what the motor primitives are and how they are combined. When dealing with motor execution, the basic question concerns the forces that are needed to produce the appropriate limb movement. Factors such as limb position, velocity, and acceleration determine the required forces (Mussa-Ivaldi and Bizzi, 2000; Poggio and Bizzi, 2004). In contrast, Chersi et al. (2006, 2010) referred to *reaching, grasping, transporting,* and *placing* as motor primitives in their proposed Chain Model for representing the separate

goal-based representations of (eating) and (placing). According to this model, pools of neurons in the mirror system, each encoding a motion primitive (e.g., *reaching for a peanut, grasping the peanut, bringing the peanut toward the mouth*) are linked together to form an overall action (e.g., *eating a peanut*). Regardless of whether motor primitives are described in terms of the kinematic variables in motor execution (limb position, velocity, etc.) or in terms of more higher-lever descriptions of motor execution (reaching for a peanut) further empirical work is needed to verify the relationship between primitives involved in motor execution and the visual recognition of actions.

For the visual recognition of actions, the visual system in most cases requires access to limb position, velocity, and acceleration. Given the tight coupling between perception and action, we would expect motor primitives for action execution and recognition to be very similar. A complementary method for determining motor primitives would be to engage participants in an action segmentation task and then assess the degree of agreement between the action kinematics and the segmentation behavior of the participants. Recent advances in understanding how humans represent events have been made by asking people to segment events (Newtson and Engquist, 1976; Newtson et al., 1977; Zacks, 2004; Zacks and Swallow, 2007; Zacks et al., 2009; Meyer et al., 2010). Zacks (2004) for instance presented participants with short animations of a circle and a square. They were then asked to segment the animations into both fine- and coarse-grained meaningful units. The results showed that movement features (e.g., position, velocity, acceleration, etc.) of the stimuli could reliably predict the event segmentation, especially for the fine-grained segmentation. In another set of experiments, Zacks et al. (2009) used more naturalistic action events. For these experiments, movies of a person folding laundry, building a house out of blocks, and assembling a video game system were segmented by participants. Movement variables associated with each event were recorded by a motion tracking system. Again, the participants were asked to segment the movies into fine-grained and coarse-grained units. The results confirmed previous findings showing that movement features were significantly correlated with segmentation behavior. For example, the speed and acceleration of body parts indicated breakpoints between action segments. This shows that the visual parsing of fairly complex events is tied to the kinematics of the stimuli.

Zacks et al. (2009) included an additional manipulation that investigated the influence of top-down conceptual knowledge vs. bottom-up driven processing on the segmentation behavior of the participants. This was done by converting the previous stimuli to only show the actor's head and hands, and their relation to one another. The idea here was that if participants have information about the meaning of the action, then they will be more inclined to segment the stimuli on the basis of that conceptual knowledge and less on the finer kinematic features, which should lead to a coarser segmentation. One group of participants (top-down) was told that they would see the same previously presented movies but this time as animations of the actor's head and hands. Another group was only told that they would see an animation of an actor performing a daily activity and that the motion of his hands and head are represented by objects used to record their movement. In this case, no contextual information about physical objects or surroundings was available to the participants. The results showed that this manipulation had no measurable effect on segmentation behavior. Regardless of previous knowledge, participants continued to base their segmentation on the perceptually salient movement features. When other sources of information were removed by just showing the simple animations, relations between movement and segmentation were strengthened. This result stands in contrast to results obtained by Castellini et al. (2007) for their model of motor execution for reaching and grasping. Prior knowledge of knowing the object (a can, a roll of duct tape, and a mug) involved in the action led to better performance for the model. Knowing the object involved in reaching and grasping allows for a more accurate hand shape in the early phase of the action.

The purpose of the present experiment is to further understand the effect of levels of processing (top-down vs. bottom-up) on the perception of movement kinematics and primitives for grasping actions in order to gain insight into possible primitives used by the mirror system as hypothesized for example by Chersi et al. (2006, 2010). Specifically, the present experiment investigates whether or not the definition of a motor primitive is affected by conceptual knowledge of the observed action. When activated knowledge about the object used in the action activates information about the possible goal of the action, this may lead to generalizations over different motor routes to the same goal. In this sense, action segmentation when guided by top-down processing may lead to fewer segments and perhaps greater agreement among people about where the relevant breakpoints between segments should be placed.

The experimental paradigm from previous studies on event segmentation (Zacks et al., 2009) will be used to investigate the role of top-down vs. bottom-up processing within the context of action segmentation for 12 grasping actions. If people are given prior information about the grasping actions being performed, this should be sufficient to activate conceptual knowledge by which to guide segmentation of the grasping actions. On the other hand, if people have no idea about the goals of the actions, then there may be a tendency to focus on more fine-grained segments of the actions because there is no information from which to abstract from small kinematic changes in the movements. Accordingly, there should be more segments and perhaps more variability about where the marks for the breakpoints should be placed for people who are only allowed to used bottom-up processing.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four participants were recruited from the student population at the University of Skövde, Sweden. Participants were randomly assigned to one of two viewing conditions. For the picture condition (top-down processing condition; $n = 12$, six females, mean age = 28) participants were first given a picture of the object that was used in each action. In the inverted + no-picture condition (bottom-up processing condition; seven females, mean age = 23) each action was mirror inverted and no-pictures of the objects were shown to the participants. Participants were also paid approximately $6.50 for their participation. All participants were informed as to the nature of the experiment including possible risks and benefits. On the basis of this information, every participant

provided written consent to their participation in the experiment. This was in accordance with Swedish law and the World Medical Association Declaration of Helsinki.

STIMULI/MATERIALS

Twelve grasping actions using one arm/hand (**Table 1**) were recorded using a ShapeHandPlus™ motion capture device (**Figure 1**). The ShapeHand™ data glove was integrated with an arm tracking ShapeTape™. This device allowed the precise capture of finger, hand, and arm positions of the person performing the actions. It also allowed the recording of actions without the objects being visible. The primary purpose of the glove was to reliably capture the kinematics of fine-grained arm and hand movements, which was not possible with other motion capture equipment such as point-light motion capture systems. For the purposes of the present experiment, the 3D coordinates of 22 reference points on the limb (the tip of each finger and thumb as well as the joints of the hand and arm) were recorded.

All 12 actions were performed by the same right-handed person using the right arm/hand to perform each action. Each action started and ended at the same resting position with the arm at the side and was performed as naturally as possible with the actual relevant objects being used in each action. Natural arm and hand actions of different durations and complexity were selected in order to sample from a wide variety of such actions and to be able to text for the sensitivity of potential differences in the number of segments for different actions. All actions, however, included moving the arm toward an object, grasping the object, using the object, releasing the object, and bringing the hand/arm down to rest again. For example, we included cyclical (or iterative) actions

such as sawing, writing, turning pages in a book, and non-cyclical actions such as opening a door, drinking from a mug, opening a can, and pouring from a bottle. We did not, however, create an *a priori* segmentation of these actions since a comparison between a predetermined number of segments and participant segmentation behavior was not the focus of the experiment. However, it was important to be able to demonstrate that the experimental method was sufficiently sensitive to detect differences in the number of segmentation marks between actions. Therefore, we included actions that seemingly had different numbers of segments. For example, *writing on a whiteboard* could be considered as consisting of the following segments: picking up a pen, taking the cap off, starting to write, underlining the text, putting the cap back on, and then putting the pen down on the table. *Opening a door* could be considered as consisting of: gripping the door handle, turning it, pushing it forward, and then releasing it. Whether or not participants would notice and mark the segments was one of the issues under investigation.

Each action was shown as a constellation of point-lights (white lights against a black background) corresponding to the reference points mentioned above. The motivation for using the point-light technique was to minimize the possible confounding of form information about the hand and arm with the kinematics of the actions. By using point-lights, participants should still be able to recognize the actions and yet not be influenced by the form information about the configuration of the hand and arm. This is certainly the case for full-body actions (e.g., Dittrich, 1993; Blake and Shiffrar, 2007) and previous results from Poinzner et al. (1981) show that point-lights attached to the fingers can reliably convey American Sign Language. There is thus reason to believe that observers should be able to recognize actions shown as point-light displays even for hand and arm actions. In addition, participants in the picture condition would be shown the objects involved in the different actions, which should facilitate recognition of the actions in a top-down manner.

Previous results using point-light displays have shown that when they are turned upside down (inverted), people have a greater difficulty recognizing the actions being performed (e.g., Dittrich, 1993; Pavlova and Sokolov, 2000; Shiffrar and Pinto, 2002). It appears that the global processing of inverted point-light displays is impaired (Hemeren, 2005), which makes action recognition difficult. Many people can see the movement of arms, legs, and hands (intact local motion processing) but have difficulty describing the movement at a higher global level of meaning (impaired global motion processing; e.g., Sumi, 1984). These results for point-light displays of biological motion are very similar to the visual processing limitations of viewing an inverted face (e.g., Carey and Diamond, 1994; Boutsen and Humphreys, 2003; Leder and Carbon, 2006). Therefore, in order to create a condition where top-down (conceptual) processing of kinematic information is severely impaired, we inverted the 12 point-light actions. It should be emphasized that despite inverting the actions the *same* kinematic information is available to the participants in both conditions. Inverting the actions does not change the kinematics present in the displays. Participants in this condition were not shown any pictures of the objects involved in the actions. All recorded action sequences were displayed in real time at 30 frames per second. Each action was also oriented to avoid as

Table 1 | List of action sequences used in the experiment (Sequence duration rounded to the nearest second).

1. Cut with scissors	(22)	7. Drink from a mug	(16)
2. Lift a dumbbell	(15)	8. Open a can and take a drink	(19)
3. Open a door	(8)	9. Solve the tower of Hanoi	(21)
4. Pour from a bottle	(12)	10. Turn pages in a book	(20)
5. Saw wood	(15)	11. Unscrew a bottle cap	(14)
6. Spray from a spray bottle	(12)	12. Write on whiteboard	(22)

FIGURE 1 | ShapeHand™ motion capture system.

much occlusion as possible. See **Figures 2 and 3** for an example of five frames from the *drink from a mug* sequence and the inverted version of the same action.

PROCEDURE

Participants were tested individually and were seated at a distance of about 60 cm to a laptop computer screen. Prior to segmenting the actions, participants were informed as to the nature of the experiment. They were told that they would see 12 brief arm-and-hand movements and that the movements would be presented as a constellation of moving point-lights. Each action was performed by one arm and hand. All participants were first instructed to simply watch an action all the way through before starting the segmentation procedure.
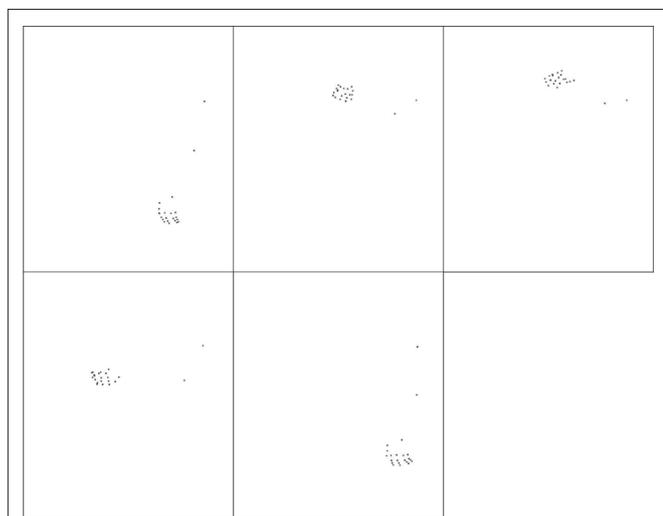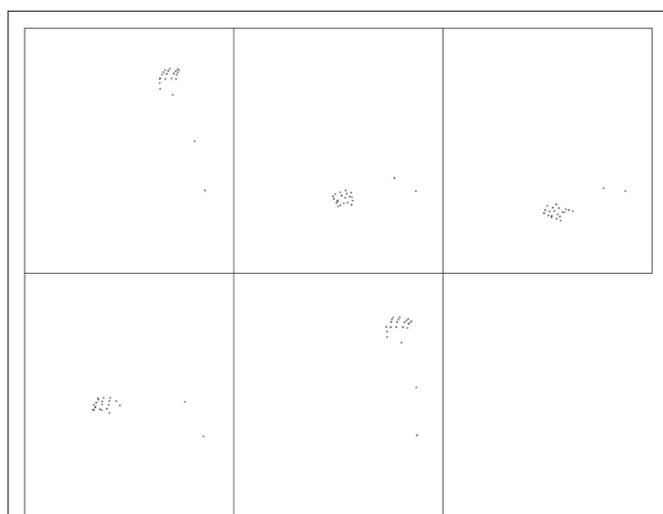
Participants in the picture condition were first presented with a picture of the object that was involved in the action, and then they viewed the action all the way through once. After viewing the action, they were asked to describe what action the person was performing and then to begin segmenting the action. Participants in the no-picture + inverted condition viewed each action without previously viewing a picture of the object used in the action. After viewing an action all the way through, the participants in this condition were instructed to begin segmenting the action without having described the action the person was performing. Following the segmentation phase, participants were asked to describe the action they thought the person was performing.

For the segmentation phase, participants were told to mark breakpoints in the action sequences that constituted the transitions between different segments in the action sequences, if they thought there were any such segments. It was left up to the participants to determine whether or not the actions contained different segments and where to mark the possible breakpoints between action segments.

Participants were also instructed how to use the stimulus player (**Figure 4**) so that they could be fairly accurate at placing marks at the breakpoints between segments. In order to reduce potential variability in participant segmentation, the stimulus player included a pause function as well as rewind and fast-forward possibilities by using the yellow slider. In other words, participants were given the opportunity to manipulate the action in order to find the desired breakpoints between segments for each action. Once they found a breakpoint between two segments, they were instructed to mark this by pressing the space bar, in which case a red line would appear on a time line below the animated figure. They were also able to make self-imposed corrections if they determined that to be necessary.

The software displaying the actions recorded the number of marks made for each action and participant as well as which frames in the action sequence were marked. There were no set time limits. The experiment lasted approximately 45 min. Participants were

FIGURE 2 | Five frames (black-on-white for clarity) at 4 s intervals from the *drink from mug* sequence, right side up.

FIGURE 3 | Five frames at 4 s intervals from the *drink from mug* sequence, inverted.
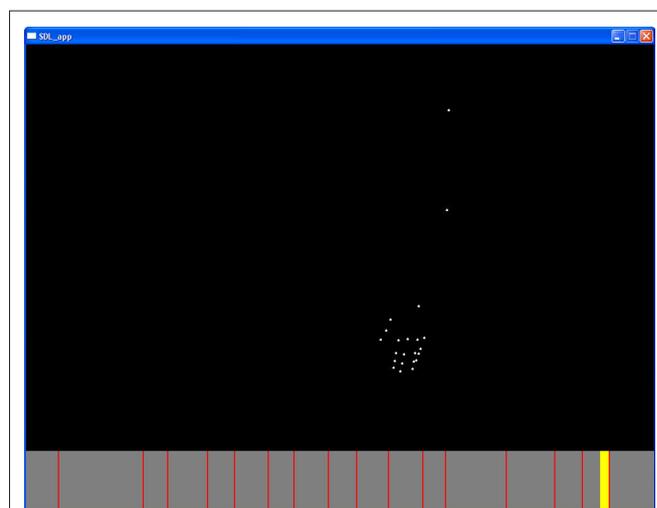
FIGURE 4 | Screen shot of an action (turning pages in a book) displayed in the stimulus player. The red lines are segmentation marks and the yellow bar is the slider.

given two practice actions to familiarize themselves with the stimuli and the tasks. The order of presentation for the 12 actions was randomly determined for each participant.

Almost all participants raised the question of how to define a segment. They were told that a segment could be viewed as a part of an action that can be used as a kind of building block to construct the whole action and that they should just mark the segments they judged to be necessary for that action. Participants had no trouble understanding the notion that actions can be segmented into action parts. The experimenter wrote down the verbal descriptions of the stimuli for later analysis. Firstly, all participant descriptions were judged for their correctness. A description was scored as correct if it included a correct identification of the action, not necessarily the object. For example, a description that included "pouring" for the *pour from bottle* action would be scored as correct. Participants' descriptions were generally clear about whether or not they recognized the actions.

In addition to participants' verbal descriptions and number of segmentation marks, the placements of the marks were analyzed in relation to the kinematic variables of velocity, acceleration, and change in direction. Velocity (measured in m/s) was obtained by approximating the first derivative of the position of recorded markers over time. It thus represents the mean tangential velocity. Acceleration (measured in m/s^2) was equivalently obtained by approximating the second derivative. The change in direction (measured in degrees/s) was computed as the angular difference between tangential vectors to the motion path at two consecutive frames in time (see **Figure 5**).

These variables were calculated for movements of the wrist as well as every finger tip during the execution of the action. To eliminate noise and minor sources of variability in the kinematic data, the obtained signals were smoothed using a standard low-pass filter (see **Figure 6**).
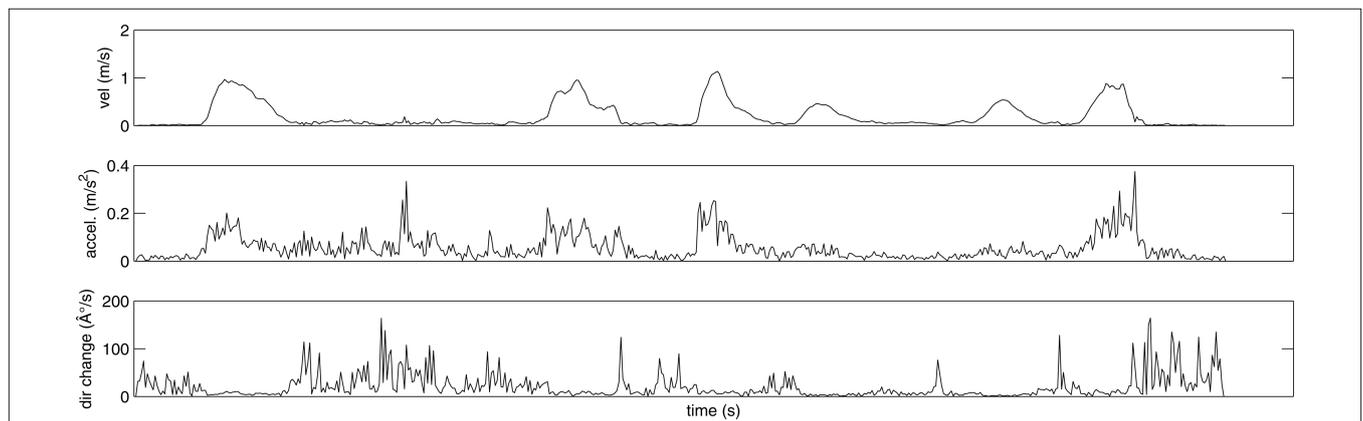
FIGURE 5 | Kinematic profile according to the time course for velocity (top), acceleration (middle), and changes in direction (bottom) for the wrist during the action *open can and drink*.
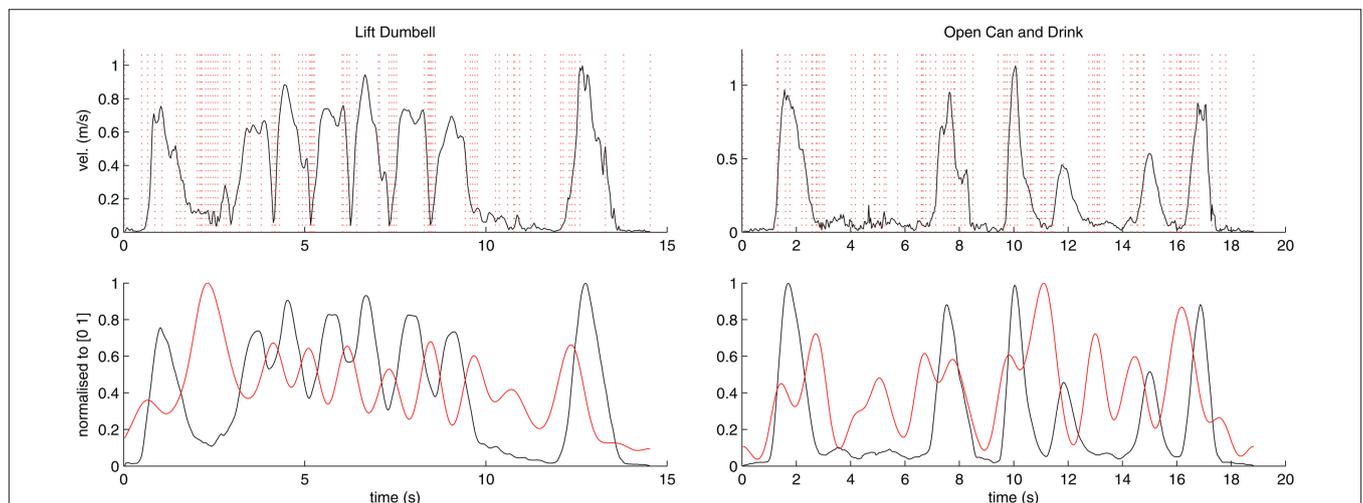
FIGURE 6 | Computation of the correlation coefficients between participant responses and kinematic data. The top row shows raw velocity data and marks placed by participants from group 1 (pictures) for two actions; *lift dumbbell* (left) and *open can and drink* (right). The bottom row shows the

corresponding velocity after smoothing through a low-pass filter (black) as well as the density function computed from the marks placed by participants (red). Curves in the bottom row are normalized to the range of [0 1] for better visualization.

RESULTS

VERBAL DESCRIPTIONS

The results for the number of correct verbal descriptions as a function of picture condition and action are presented in **Table 2**. The results show clearly that participants who viewed that actions upright and together with a picture could identify the actions presented in the point-light displays. These participants could identify the immediate goal of the action. Here are some examples of the verbal responses:

– *The person is gripping the door handle and opening the door.*
– *The person is unscrewing the cap on the bottle.*
– *The person is pouring something from the bottle.*
– *A person opened a can and then lifted the can and took a drink.*

In contrast, the participants who saw the actions upside down and did not get to view the object used in the action were impaired at recognizing the goal, or higher-level purpose, of the actions. Here are some examples of the verbal responses for participants in that condition (with the displayed action in parentheses):

– *Someone sticking their thumb out and pointing in a certain direction. (opening a door)*
– *Moves the hand down, and the thumb switches place, the thumb and index finger, and the middle finger are touching each other. (unscrewing a bottle cap)*
– *Grabs something without using the index finger. (pouring from a bottle)*

– *First phase, the hand and fingertips do something, the second phase, grabs something and holds it close and then pushes it away. (open a can and drink)*

These descriptions are typical of the participant responses in this condition and show that participants are able to see and describe the motion of the fingers and arm but fail to interpret the motions according to any higher-level action description. It appears then that participants have visual access to the kinematics of the parts of the hand and arm.

NUMBER OF SEGMENTATION MARKS

The mean number of marks as a function of viewing condition and action are presented in **Table 3**. Regarding the differences between the number of marks for viewing condition (picture vs. no-picture-inverted), participants made slightly more marks in the picture condition than in the no-picture-inverted condition. A mixed ANOVA, 2 (viewing condition: picture vs. no-picture-inverted, between subjects) × 12 (action, within subjects) on the marking data, however, showed that the main effect of viewing condition was not significant, $F < 1$. The main effect of action, however, was significant, Greenhouse–Geisser adjusted $F(4.2, 92.32) = 6.51$, $p < 0.001$. This indicates that participants viewed different actions as consisting of different numbers of segments. The interaction was not significant, Greenhouse–Geisser adjusted $F(4.2, 92.32) = 1.53$, $p = 0.20$.

Regarding the main effect of the action variable, some actions received more segmentation marks than others. For example, *post hoc* comparisons showed that participants marked significantly more segments for *writing on the board* ($M = 7.7$) compared to

Table 2 | Number of correct verbal descriptions of the actions.

	Action					
Viewing condition	Cut with scissors	Lift dumbbell	Open door	Pour from bottle	Saw wood	Spray with spray bottle
Pictures	12	12	12	12	12	12
No-pictures inverted	4	1	1	2	2	1
	Drink from mug	Open a can and drink	Move disks tower Hanoi	Turn pages in book	Unscrew bottle cap	Write on board
Pictures	12	12	12	12	11	12
No-pictures inverted	1	1	1	0	1	3

Table 3 | Mean number of segmentation marks as a function of viewing condition and action (SDs in parentheses).

	Action					
Viewing condition	Cut with scissors	Lift dumbbell	Open door	Pour from bottle	Saw wood	Spray with spray bottle
Pictures	7.7 (5.6)	7.8 (3.7)	4.7 (2.1)	5.0 (2.0)	9.7 (6.9)	6.7 (3.0)
No-pictures inverted	6.5 (4.7)	6.5 (4.3)	4.3 (1.9)	5.5 (2.4)	6.0 (3.3)	6.5 (3.6)
	Drink from mug	Open a can and drink	Move disks tower Hanoi	Turn pages in book	Unscrew bottle cap	Write on board
Pictures	5.6 (1.8)	7.8 (2.7)	12.0 (7.9)	9.2 (6.0)	6.0 (3.8)	8.3 (5.3)
No-pictures inverted	5.3 (2.1)	6.9 (2.5)	7.6 (6.8)	7.8 (4.5)	5.8 (3.3)	7.0 (2.9)

opening a door $(M = 4.5)$, $t(22) = 4.66$, $p = 0.008$. Other Bonferroni adjusted *post hoc* comparisons showed that *lifting a dumbbell* $(M = 7.1)$, *opening a can and drinking* $(M = 7.4)$, *move disks on the Tower of Hanoi* $(M = 9.8)$, and *turning pages in a book* $(M = 8.5)$ had significantly more segment marks than *opening a door*, ps < 0.05. *Opening a can and drinking* also had significantly more segmentation marks than *pouring from a bottle* $(M = 5.3)$ and *drinking from a mug* $(M = 5.4)$, $p < 0.05$. These differences will be further discussed in the Section "Discussion."

The large standard deviations for the segmentation marks for some of the conditions suggests that participants the number of segmentation marks varied quite a bit for some of the actions. This seems to be a result of a difference in marking behavior for the repetitive actions like *cutting with scissors, sawing wood, moving disks tower of Hanoi, turning pages in a book,* and *writing on a board*. This will also be discussed in the Section "Discussion."

CORRELATIONS BETWEEN KINEMATIC VARIABLES AND SEGMENTATION

For every action, a density function of the marks placed along the timeline by each group of participants was computed using a Gaussian kernel with a width of 0.3 s (see **Figure 6**). Linear (Pearson) correlation coefficients between the density function and each of the kinematic variables (velocity, acceleration, and change in direction) were then calculated (**Table 4**). Although kinematic variables were tracked for the wrist as well as for every finger, we did not in general find any noteworthy differences in the correlation coefficients between the different points on the hand and wrist and the density functions. For simplicity, we therefore limit the discussion in this paper to the behavior of the wrist kinematics. Varying the width of the kernel within reasonable limits (i.e., avoiding extreme cases in which the peaks of the density function are exceedingly narrow or in which the density function is so smeared out that details of the marking behavior are lost) did not generally affect whether or not correlations were significant.

The results in **Table 4** show that change in direction is significantly inversely correlated with 10 of the 12 actions in the picture condition and with 11 of the 12 actions in the no-picture-inverted condition, indicating that change in direction is associated with fewer segmentation marks. Segmentation marking is more positively associated with other kinematic variables. Velocity, for example, is significantly positively correlated with seven actions in the

Table 4 | Correlations (Pearson r) between mark density function and kinematic variables measured from the wrist: change in direction of movement, velocity, and acceleration.

| | Viewing condition: with pictures | | | | | |
| | | | Action | | | |
Wrist	**Cut with scissors**	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Change direction	−0.18*	−0.29*	−0.34*	−0.35*	−0.48*	0.04
Velocity	0.12*	0.03	0.02	−0.01	0.43*	−0.02
Acceleration	0.12*	0.29*	0.31*	0.11	0.42*	0.09
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Change direction	−0.22*	−0.28*	−0.26*	−0.12*	−0.02	−0.45*
Velocity	0.33*	0.18*	0.01	0.19*	0.37*	0.44*
Acceleration	0.23*	0.20*	0.06	0.35*	0.41*	0.34*
	Viewing condition: no-picture-inverted					
			Action			
	Cut with scissors	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Change direction	−0.49*!	−0.11	−0.33*	−0.47*	−0.52*	−0.40*!
Velocity	0.46*!	−0.14*	0.45*!	0.38*!	−0.13*	0.51*!
Acceleration	0.32*!	0.20*	0.75*!	−0.41*!	−0.40*	0.28*!
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Change direction	−0.22*	−0.32*	−0.36*	−0.33*!	−0.15*	−0.54*
Velocity	0.53*!	0.36*!	0.48*!	0.46*!	0.54*!	0.51*
Acceleration	0.50*!	0.29*	0.44*!	0.40*	0.53*	0.30*

A (*) indicates a significant coefficient at the 0.01 level. Statistically significant coefficients are in bold-face type. A (!) indicates a significant difference between that coefficient and the coefficient in the corresponding condition for the viewing condition with pictures at the 0.01 level.

picture condition and with 10 actions in the no-picture-inverted condition, which indicates that segmentation marking increases with velocity. Acceleration is significantly positively correlated with nine actions in the picture condition and with 10 actions in the no-picture-inverted condition.

Higher values of velocity appear to signal the start of a segment whereas an episode of ongoing changes in direction tends to be associated with the carrying out of a part of an action. For example, for the action of *drinking from a mug*, there are changes in direction during the drinking phase which consists of tilting the mug and consequently thereby changing the direction of the wrist point. The act of drinking as such appears to be a whole segment and is not further segmented. An analysis of this inverse relationship between change in direction and velocity (**Table 5**) shows that it is significant for all actions. This effect indicates that it is not the occurrence of a change in direction for the wrist that is associated with action segmentation.

The size of many of the correlation coefficients for the no-picture-inverted group are slightly higher than the corresponding coefficients in the picture group, which indicates that the segmentation behavior of the participants in the no-picture-inverted group is more strongly related to the kinematics for those correlations.

Comparisons (z-transformed) between the correlation coefficients for the two viewing conditions, actions, and kinematic variables showed that 18 of the 36 coefficients for the no-picture-inverted group were significantly larger (assuming the same direction of correlation) than the corresponding coefficients for the picture group. This finding will be further in the Section "Discussion."

AGREEMENT ACROSS CONDITIONS

Despite some differences between the picture and the no-picture-inverted conditions, participants in both groups appear to be marking similar segments. Correlation coefficients for the relationship between the density functions for the picture and no-picture-inverted groups show significant agreement (**Table 6**). The range between the highest (unscrew a bottle cap) and lowest (tower of Hanoi) correlation coefficient is quite large and suggests different levels of agreement for the different actions. Plots of the density functions for the segmentation marks for the two groups and for two actions (*drink from mug* and *tower of Hanoi*) are presented in **Figure 7**.

Table 5 | Correlation coefficients (Pearson *r*) for the relationship between change in direction and velocity for the actions.

Action	*r*	Action	*r*
1. Cut with scissors	−0.23	7. Drink from a mug	−0.45
2. Lift a dumbbell	−0.51	8. Open a can and take a drink	−0.23
3. Open a door	−0.45	9. Solve the tower of Hanoi	−0.21
4. Pour from a bottle	−0.30	10. Turn pages in a book	−0.34
5. Saw wood	−0.50	11. Unscrew a bottle cap	−0.30
6. Spray from a spray bottle	−0.32	12. Write on whiteboard	−0.39

All coefficients are significant at the 0.001 level.

Table 6 | Correlation coefficients (Pearson *r*) for the relationship between the density functions for the picture and no-picture-inverted groups.

Action	*r* (*n*)	Action	*r* (*n*)
1. Cut with scissors	0.72 (650)	7. Drink from a mug	0.86 (477)
2. Lift a dumbbell	0.71 (435)	8. Open a can and take a drink	0.50 (564)
3. Open a door	0.60 (236)	9. Solve the tower of Hanoi	0.18 (634)
4. Pour from a bottle	0.60 (343)	10. Turn pages in a book	0.58 (584)
5. Saw wood	0.46 (457)	11. Unscrew a bottle cap	0.93 (404)
6. Spray from a spray bottle	0.66 (349)	12. Write on whiteboard	0.67 (670)

All coefficients are significant at the 0.001 level (n = number of observations of velocity and mark density function over the time course of the action, i.e., number of frames for each action).
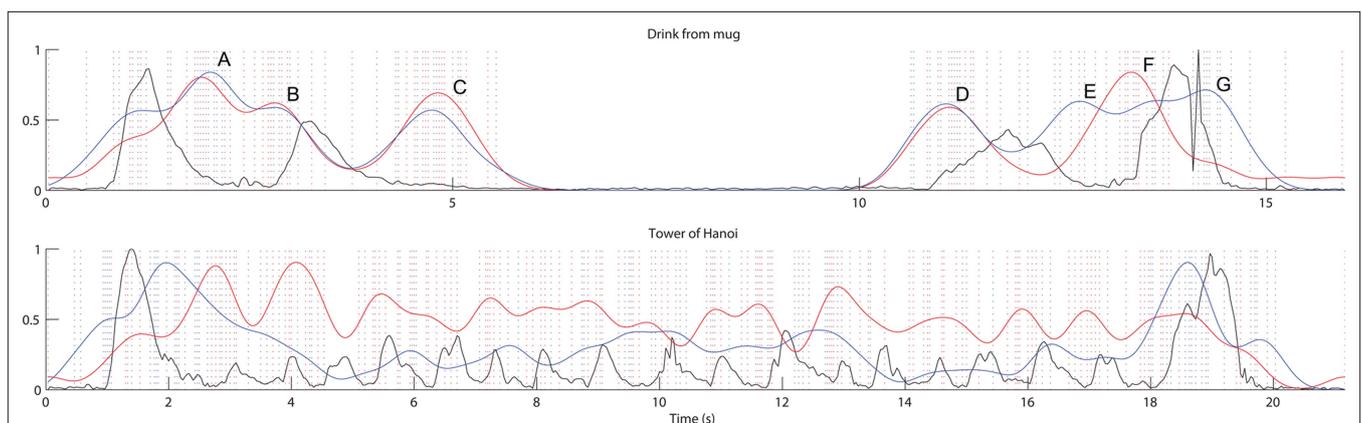
FIGURE 7 | Relationship between density functions for the segmentation marks for the picture and no-picture-inverted groups for two actions.
Black = velocity. Red = picture group. Blue = no-picture group. Dotted vertical lines indicate marks placed by participants in the picture (red) and no-picture group (blue) respectively.

The top density function profile in **Figure 7** shows the considerable agreement between participants for the two different groups for the action *drinking from a mug*. For this action (top figure in **Figure 7**), participants marked gripping the mug (A), starting to lift the mug (B), starting to drink (C), stop drinking (D), putting the mug down, blue line (E), releasing the grip, red line (F), and setting the mug down again (G). In this case there was a small difference between the groups regarding the marking of release of the grip. The picture group appeared more inclined to mark that segment than the no-picture-inverted group.

For the bottom density function profile (tower of Hanoi), the main difference between the groups seems to concern whether or not to mark the recurring grasping–moving–releasing motions involved in the action. These differences will be further discussed in the Section "Discussion" below.

DISCUSSION AND CONCLUSION

In response to the central question of the experiment reported here, the results show that there is a significant relationship between segmentation behavior and velocity or acceleration and change of direction for most of the hand/arm actions. This result holds for the picture group and the no-picture-inverted group, which suggests that differences between top-down and bottom-up visual processing of the point-light hand/arm actions did not carry over to obvious differences in overall segmentation behavior. This conclusion is also supported by the significant correlations between the density profiles for the two different conditions (**Figure 7**), which indicate that participants in the two groups tend to place their segmentation marks in similar locations. The segmentation behavior of participants suggests that an increase in velocity (or acceleration for the picture group) is an important signal for denoting a segment boundary, or breakpoint.

There were also no significant differences between the number of segmentation marks for the picture and no-picture-inverted conditions (**Table 3**), which is consistent with the conclusion that the two groups of participants are similar in their segmentation behavior. Admittedly, drawing conclusions on the basis of a null-effect is somewhat problematic in that it can be difficult to determine whether the null-effect is the result of an insufficiently sensitive method or the result of there actually being no effect of the independent variable. With regard to the issue of an insufficiently sensitive method, there were significant differences between segmentation marks for the different actions, which suggests that the sensitivity of the method was sufficiently high to also detect potential differences between the different viewing conditions. There were, however, some rather high standard deviations for a few of the conditions. We purposely included a number of different actions in order to detect possible differences between different kinds of actions. This was done because there is no previous research that has investigated the visual segmentation of hand/arm actions. By including many different actions, the results could potentially have a greater external validity than if the segmentation stimuli were limited to fewer actions.

A similar line of reasoning can be applied to the freedom that was given to participants for the task of placing segmentation marks, which also likely contributed to the variability. Participants were given the possibility of moving the yellow slider (**Figure 4**) in order to find the frame they wanted to mark as a breakpoint between segments, and all participants used the slider to try and find the breakpoints between segments, but they realized it took far too much time to exactly specify the frame for each breakpoint for 12 actions. We did not require exact precision in the segment marking task because we wanted to avoid demand characteristics associated with too many constraints on the segmentation task. The smoothing function mentioned above was used to treat this variation and has been used by other researchers on similar data (Meyer et al., 2010), or alternatively, a binning technique has been used (e.g., Zacks, 2004).

VERBAL DESCRIPTIONS AND RECOGNITION

The results from the verbal descriptions show that the orientation manipulation for the point-light hand/arm actions successfully created top-down and bottom-up driven visual processing of the stimuli. This is consistent with the findings from previous studies of whole body point-light displays (e.g., Bertenthal and Pinto, 1994; Pavlova and Sokolov, 2000). Recognition of the actions was obviously impaired by inverting the action. Despite the severe impairment of not being able to recognize the actions, participants were still able to consistently use the kinematics to mark segments of the actions, which was indicated by the significant correlations with change in direction, velocity, and acceleration. Furthermore, the segmentation behavior of participants who viewed the inverted displays was reliably correlated with the segmentation behavior of the participants who viewed the actions in an upright orientation. This shows that both groups seem to base their segmentation of the actions on the more low-level kinematics rather than high-level knowledge associated with the conceptual understanding of the viewed actions. The participants in the no-picture-inverted condition seem to do this a bit more. Top-down influences do not seem to modulate segmentation behavior such that they lead to very different segmentation behavior. Our results are consistent with the results from Zacks et al. (2009) and Bidet-Ildei et al. (2006). In the results from Bidet-Ildei et al. (2006) participants were able to reliably discriminate between natural and unnatural arm movements of point-light displays of elliptic motion but were very poor at identifying the display as a specific arm movement of a human making an elliptic motion. Results from our experiment show that the inversion manipulation seems to work on even smaller grained actions (more local limb motions). There appears to be local motion processing of the limb parts, e.g., fingers and hand, and more global (holistic) processing seems to be impaired.

The fact that participants can see and describe the movements of body parts but fail to identify the higher-level semantic meaning of the actions is similar to association agnosia for objects where patients can see the parts of objects but fail to identify the object as such (Farah, 2004). There is no strictly visual deficit as such but rather an inability to recognize the object. When our participants view the inverted point-light actions, they are able to visually discern the relevant body parts and segment the actions on the basis of changes in the direction of movement and velocity. What seems to be missing is the epistemic visual perception (Jeannerod and Jacob, 2005).

POSSIBLE INFLUENCE OF TOP-DOWN PROCESSING

One difference mentioned above between the two viewing conditions is that some of the correlation coefficients were somewhat higher for the no-picture-inverted group. Although segmentation marking appears to be similar for the two conditions, participants in the no-picture-inverted condition tend to base their segmentation marks somewhat more on the kinematics than the participants in the picture group. One explanation for this behavior could be that participants with access to higher-level meaning of actions tend to rely less on the precise pattern of the kinematics during action segmentation. The results from the verbal descriptions seem to confirm this as participants with access to higher-level information tend to describe segmentation instances based on this [e.g., "reached out his hand, gripped the (door) handle and opened the door"] whereas the second group (no-picture-inverted) tended to refer more closely to actual movements (e.g., "grips something from underneath and turns it"). One such example of where top-down information appears to influence the segment marking can be seen in the case of *opening a door*. Toward the end of that action, participants who were shown a picture of a door handle clearly marked the releasing of the grip on the door handle, i.e., segmentation was about the grip. For the group that saw the action upside down and without a picture, marks were made in connection with the motion of the hand downward toward a resting position. The major difference here is that for the picture group, the segmentation was about the grip, and for the no-picture group the segmentation was about the movement of the arm/hand. This suggests that top-down knowledge seems to involve gripping as an important segment for opening a door, which is not similarly marked for the group that did not have access to any top-down knowledge about the action. Future experiments will have to look more closely at this potential difference.

Another possibility for a potential role of high-level knowledge concerns action hierarchies. Poggio and Bizzi (2004) discuss the combination of motor primitives into hierarchical representations that allow generalization over specific situations. If participants are given the task of segmenting inverted and upright actions on the basis of very coarse-grained patterns and not allowed to place as many breakpoints as they wish, we should see more pronounced differences between top-down and bottom-up processing.

We are thus not making the strong claim that there is no modulation of the segmentation of action sequences by top-down visual processing. It could be case that given other manipulations, epistemic visual perception may lead to different segmentation strategies. However, our results regarding the overall segmentation behavior do indicate that the different strategies may nonetheless converge onto similar marking behavior.

MOTOR PRIMITIVES AND THE MIRROR SYSTEM

As discussed in the introduction, it is still an open question regarding what motion primitives are encoded by the mirror system. The chain model (Chersi et al., 2006) is primarily based on neurophysiological data from Macaque monkeys (e.g., Fogassi et al., 2005) and therefore includes only very basic primitives that correspond to easy tasks that monkeys can carry out in laboratory settings. It is an open question how well, if at all, those findings can translate to a human mirror system and therefore, it is of interest to identify potential "human" motion primitives.

Since it is at present almost impossible to record from relevant human neurons at an adequate resolution to answer this question, we have chosen an action segmentation task instead. Critically, while the segmentation behavior of the no-picture-inverted (bottom-up) group in several instances correlated more closely with the kinematics than the picture group, our results indicate that the overall segmentation behavior nonetheless remains similar; it thus seems that primitives which are identifiable in a task such as the one in the present paper are defined primarily by the kinematics of the actions rather than additional contextual information, although the latter can clearly influence the identification of start and end points of primitives.

This is interesting in the light of the argument that mirror neurons encode not only a motion primitive but also the goal of the action an observed or executed primitive is part of, e.g., Fogassi et al. (2005), Umiltà et al. (2008). One could infer from such an entanglement of encodings that higher-level contextual information affects the definition of primitives. However, our results indicate that this is not the case.

It should also be noted here that kinematics are not necessarily equal to motor commands or muscle activations. Umiltà et al. (2008) for instance have shown that mirror neurons in monkeys do not encode the motor commands needed to execute an action but rather the behavior of the end-effector, and the role of kinematic variables in action segmentation studied here reflect that aspect.

CONCLUSION

When participants are given the task of segmenting hand/arm actions presented as point-light displays, segmentation is largely based on the kinematics, i.e., the velocity and acceleration of the wrist and hand extremities (finger tips), regardless of whether or not participants have access to higher-level information about the action. If access to high-level information about the identification of the action, e.g., drinking) is impaired by inverting the point-light displays, the kinematic information remains a salient source of information on which to base action segmentation. If participants have access to the high-level information, they still tend to rely on the kinematics of the hand/arm actions for determining where to place segmentation marks. This indicates that top-down activation of an action representation leads to similar segmentation behavior for hand/arm actions compared to bottom-up, or local, visual processing when performing a fairly unconstrained segmentation task. Future studies will need to address the issues of more precisely identifying motor primitives and determining their hierarchical organization in relation to high-level knowledge structures.

SUPPLEMENTARY MATERIAL

The Movies 1 and 2 for this article can be found online at http://www.frontiersin.org/cognition/10.3389/fpsyg.2010.00243/abstract

REFERENCES

Bertenthal, B. I., and Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5, 221–225.

Bidet-Ildei, C., Orliaguet, J.-P., Sokolov, A. N., and Pavlova, M. (2006). Perception of elliptic biological motion. *Perception* 35, 1137–1147.

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73.

Blakemore, S.-J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567.

Borghi, A. M., and Riggio, L. (2009). Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Res.* 1253, 117–128.

Boutsen, L., and Humphreys, G. W. (2003). The effect of inversion on the encoding of normal and "thacherized" faces. *Q. J. Exp. Psychol.* 56A, 955–975.

Buccino, G., Binkofski, F., and Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376.

Carey, S., and Diamond, R. (1994). Are faces conceived as configurations more by adults than by children? *Vis. Cogn.* 1, 253–275.

Castellini, C., Orabona, F., Metta, G., and Sandini, G. (2007). Internal models of reaching and grasping. *Adv. Robot.* 21, 1545–1564.

Chersi, F., Mukovskiy, A., Fogassi, L., Ferrari, P. F., and Erlhagen, W. (2006). A model of intention understanding based on learned chains of motor acts in the parietal lobe. *Comput. Neurosci.* 69, 48.

Chersi, F., Thill, S., Ziemke, T., and Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Front. Neurorobotics* 4:4. doi: 10.3389/fnbot.2010.00004

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events. *Exp. Brain Res.* 91, 176–180.

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

Farah, M. J. (2004). *Visual Agnosia*. Cambridge, MA: MIT Press.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667.

Giese, M. A., and Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Res.* 38, 1847–1858.

Grafton, S. T., Fadiga, L., Arbib, M. A., and Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* 6, 231–236.

Hemeren, P. E. (2005). "Orientation specific effects of automatic access to categorical information in biological motion perception," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, eds B. G. Bara, L. W. Barsalou, and M. Bucciarelli (Hillsdale, NJ: Erlbaum), 935–940.

Hemeren, P. E. (2008). *Mind in Action: Action Representation and the Perception of Biological Motion*. Lund University Cognitive Studies 140. Sweden: Lund University.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3, e79. doi: 10.1371/journal.pbio.0030079

Jeannerod, M. (1995). Mental imagery in the motor cortex. *Neuropsychologia* 33, 1419–1432.

Jeannerod, M., and Jacob, P. (2005). Visual cognition: a new look at the two-visual systems model. *Neuropsychologia* 43, 301–312.

Leder, H., and Carbon, C.-C. (2006). Face-specific configural processing of relational information. *Br. J. Psychol.* 97, 19–29.

Meyer, M., DeCamp, P., Hard, B., Baldwin, D., and Roy, D. (2010). "Assessing behavioral and computational approaches to naturalistic action segmentation," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 2710–2715.

Michelon, P., Vettel, J. M., and Zacks, J. M. (2006). Lateral somatotopic organization during imagined and prepared movements. *J. Neurophysiol.* 95, 811–822.

Mussa-Ivaldi, F. A., and Bizzi, E. (2000). Motor learning through the combination of primitives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1755–1769.

Newtson, D., and Engquist, G. (1976). The perceptual organization of ongoing behavior. *J. Exp. Soc. Psychol.* 12, 436–450.

Newtson, D., Engquist, G., and Bois, J. (1977). The objective basis of behavior units. *J. Pers. Soc. Psychol.* 35, 847–862.

Pavlova, M., and Sokolov, A. (2000). Orientation specificity in biological motion perception. *Percept. Psychophys.* 62, 889–899.

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. *Nature* 431, 768–774.

Poinzner, H., Bellugi, U., and Lutes-Driscoll, V. (1981). Perception of American sign language in dynamic point-light displays. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 430–440.

Pollick, F. E. (2004). "The features people use to recognize human movement style," in *Gesture-Based Communication in Human–Computer Interaction*, Vol. 2915, eds A. Camurri and G. Volpe (Berlin: Springer-Verlag), 10–19.

Pollick, F. E., Fidopiastis, C., and Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception* 30, 323–338.

Pozzo, T., Papaxanthis, C., Petit, J. L., Schweighofer, N., and Stucchi, N. (2006). Kinematic features of movement tunes perception and action coupling. *Behav. Brain Res.* 169, 75–82.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2006). Mirrors in the mind. *Sci. Am.* 295, 54–61.

Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D. I., and Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.

Shiffrar, M., and Pinto, J. (2002). "The visual analysis of bodily motion," in *Common Mechanisms in Perception and Action: Attention and Performance*, Vol. 19, eds W. Prinz and B. Hommel (Oxford: Oxford University Press), 381–399.

Sumi, S. (1984). Upside-down presentation of the Johansson moving light pattern. *Perception* 13, 283–286.

Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747.

Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., and Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2209–2213.

van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cogn. Sci.* 28, 979–1008.

Zacks, J. M., Kumar, S., Abrams, R. A., and Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition* 112, 201–216.

Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Curr. Dir. Psychol. Sci.* 16, 80–84.