Communicative and noncommunicative point-light actions featuring high-resolution representation of the hands and fingers

**Hazlin Zaini · Jonathan M. Fawcett ·
Nicole C. White · Aaron J. Newman**

Abstract We describe the creation of a set of point-light movies depicting 43 communicative gestures and 43 noncommunicative, pantomimed actions. These actions were recorded using a motion capture system that is worn on the body and provides accurate capture of the positions and movements of individual fingers. The movies created thus include point-lights on the fingers, allowing for representation of actions and gestures that would not be possible with a conventional, line-of-sight-based motion capture system. These videos would be suitable for use in cognitive and cognitive neuroscientific studies of biological motion and gesture perception. Each video is described, along with an *H* statistic indicating the consistency of the descriptive labels that 20 observers gave to the actions. We also produced a scrambled version of each movie, in which the starting position of each point was randomized but its local motion vector was preserved. These scrambled movies would be suitable for use as control stimuli in experimental studies. As supplementary materials, we provide QuickTime movie files of each action, along with text files specifying the three-dimensional coordinates of each point-light in each frame of each movie.

Keywords Biological motion · Motion capture · Point-light · Gesture · Pantomime · Stimuli

H. Zaini · J. M. Fawcett · N. C. White · A. J. Newman (✉)
Department of Psychology & Neuroscience, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
e-mail: aaron.newman@dal.ca

The term "biological motion" refers to motion patterns generated by the actions of animals and includes acts such as walking or gesturing. This topic was first formally investigated by Johansson (1973), who discovered that human motion was easily identifiable from a few (8–12) points of light on the body. Since that time, it has been shown that rich information about a person can be gleaned from point-light biological motion, including the actor's sex, mood, and identity, as well as the actions being performed (e.g., Barclay, Cutting, & Kozlowski, 1978; Cutting & Kozlowski, 1977; Dittrich, 1993). Point-light stimuli are particularly useful because they isolate the minimal information necessary for human action recognition: the coordinated, global movement of the parts of the body relative to each other and to gravity. Inverting point-light actions disrupts recognition (Dittrich, 1993), but when all points except the feet are inverted, people can still identify the direction of walking (Troje & Westhoff, 2006). Thus, point-light stimuli allow for strong experimental control by allowing us to isolate biological motion from extraneous information such as the appearance of the actor, facial expressions, and so forth. In neuroimaging experiments, this can be extremely valuable, since techniques such as fMRI typically rely on a subtraction methodology. In this method, the brain activation of interest is isolated from activation associated with other factors by comparing the signals between two conditions that differ only in the presence or absence of the factor of interest. For example, biological motion has been isolated by comparing point-light movies of human actions with movies in which the starting point of each point-light was randomized but the motion vectors over time were preserved (Grossman et al., 2000).

Although much research has focused on point-light walkers (e.g., Giese & Poggio, 2003), stimulus sets have been developed and made publicly available depicting a variety of common activities (e.g., chopping wood or dancing) as point-light movies (Dekeyser, Verfaillie, & Vanrie, 2002; Ma, Paterson, & Pollick, 2006; Vanrie & Verfaillie, 2004). Such standardized stimulus sets are valuable in providing the research community with stimuli that are standardized, accessible, and associated with normative data. Indeed, we have used Vanrie and Verfaillie's stimuli in an event-related brain potential study of biological motion (White, Fawcett, & Newman, 2009). In our neuroimaging work on the neural substrates of gesture and sign language comprehension, we have found a need for point-light versions of such communicative actions. Achieving a "tight" subtraction for neuroimaging studies of American Sign Language (ASL) has proved challenging via standard video-recording methods. We have used nonsense signing (Newman, Bavelier, Corina, Jezzard, & Neville, 2002); however, it was impossible to ensure that the actor produced such strings of actions with natural fluency, prosody, or facial expression, since the signs are devoid of the meaning that is associated with communicative signals. In more recent work, we initially used ASL sentences played in reverse, but we found that ASL signers were able to understand these. Thus, we further overlaid three reversed sentences semitransparently (Newman, Supalla, Hauser, Newport, & Bavelier, 2010a, 2010b), which disrupted linguistic comprehension. Since the same movies were used in "normal" and "backward-overlaid" signing, both types of stimuli contained the same information (aggregated over the entire stimulus set); however, the backward-overlaid signing contained more of all of the nonlinguistic information (biological motion, faces, etc.) in the original signals. While these stimuli proved effective in isolating brain activation related to sign language and also to nonlinguistic gesture (Newman, Newport, Supalla, & Bavelier, 2007), point-light stimuli offer the benefit of even tighter experimental control.

In attempting to create such stimuli, we initially tried different systems, including those using infrared emitters attached to the body, high-speed video cameras, and magnetic "points" attached to the body. These all suffered from limitations. First, being able to capture gestures and sign language requires resolution of the independent movements of each finger on each hand. Emitter-based systems required large numbers of emitters to be taped to both the inside and outside surfaces of the fingers and hands, which was time consuming and particularly awkward for the actor, who had to contend with more that two dozen wires coming off each hand. Magnetic systems similarly required impractically large numbers of sensors. Second, any optical system suffered from line-of-sight limitations, whereby point tracking failed whenever one of the emitters or reflective points was occluded as a result of the hand or arm turning or blocking the other hand/arm. While creating stimuli with optical equipment would have been technically possible, the number of cameras required and their orientations would have been prohibitively expensive. Older studies of point-light motion, including of sign language (Tartter & Fischer, 1982), had used white or reflective tape filmed in a dark room. While effective, these methods do not easily lend themselves to digitization, and thus to the power and flexibility of having three-dimensional coordinates of each point over time.

In the present work, we used a motion capture system composed of flexible, fiber-optic bend-and-twist sensors, combined with accelerometers and magnetometers, that is worn on the body. This system allowed for accurate capture of the individual fingers along with the rest of the upper body. We used this system to develop a set of point-light action movies that supplements and extends the extant point-light biological motion stimulus banks. Half of the stimuli were instrumental, pantomimed actions, including some used in previous stimulus sets (Dekeyser et al., 2002; Ma et al., 2006; Vanrie & Verfaillie, 2004). They complement other available sets because of the presence of point-lights on the hands and fingers, however. The other half of the stimuli were communicative gestures, of the type commonly referred to as "emblems"—actions that have a commonly understood and agreed-upon meanings within a culture/language group and can typically stand for a word or phrase (Ekman & Friesen, 1969; McNeill, 1985). These are, to our knowledge, unique among readily available stimulus sets.

Here we describe the creation of these stimuli, as well as the process of selecting the final stimuli on the basis of the labels assigned to them by 20 naïve observers. The stimuli are provided in the supplementary materials, as both video files and text files specifying the location of each dot in three-dimensional space across time. It is important to note that our intention in developing and releasing these materials was that other researchers might find them useful as stimuli in studies of biological motion and gesture processing. Our primary goal was to ensure that the actions depicted in the point-light movies would be readily recognizable by viewers, rather than to preserve the kinematic accuracy of the original recorded movements.

Stimulus creation

Motion capture hardware

Biological motion was recorded using a wearable, wireless motion capture system (ShapeWrap III) developed by

Measurand Inc. (Fredericton, NB, Canada). This system, shown in Fig. 1, consisted of sensors for the upper body, including (a) a head orientation sensor, (b) a thoracic orientation sensor, (c) a pelvic orientation sensor, (d) two arm sensors, (e) four finger sensors for each hand, and (f) thumb sensors for each hand. All of the sensors were tethered to (g) a data concentrator located on the back of the upper body. Motion capture using this hardware was not based on the optical capture of "markers" placed at specific points on the body, nor on a set of magnets, as in many other systems, including those used in the development of previous biological-motion stimulus sets (Dekeyser et al., 2002; Ma et al., 2006; Vanrie & Verfaillie, 2004). Rather, data were combined from a set of inertial/orientation sensors and a set of fiber-optic "bend-and-twist" tapes. This is important, because many of our stimuli involved subtle articulation of the hands or fingers, as well as frequent rotation of the wrist that would periodically occlude various body surfaces throughout recording. Optical systems depend on a line of sight, which is frequently occluded during gestures and actions such as those that we used, making it impossible to preserve all markers throughout the movements. Most optical and magnetic systems rely on a set number of markers placed on the body. This can limit resolution, as well as the actor's comfort and/or freedom of movement, which may in turn disrupt natural movement patterns.

The inertial/orientation sensors were composed of tri-axial accelerometers, magnetometers, and angular-rate sensors. The angular-rate sensors measured angular rotations about each x-, y-, and z-axis. To account for the drift inherent in these measures, each inertial/orientation sensor also included an accelerometer and a magnetometer to correct for this drift within each axis. The accelerometers measured tilt, and the magnetometers used the Earth's magnetic field to measure direction (like a compass). The pelvic orientation sensor measured the actor's pelvis in terms of a world coordinate system (WCS), while the thoracic orientation sensor and the head orientation sensor measured the orientation of the actor's torso and head, respectively, in terms of a body coordinate system (BCS). The x-axis of the BCS faces forward from the pelvis, the y-axis points from the center of the pelvis toward the head, and the z-axis points from the center of the pelvis toward the right hip. Most joint angles and positions were reported relative to the WCS. The origin of the WCS can be considered as an imaginary point in the center of the floor within the recording software. The x-axis of the WCS is "forward," the y-axis is "up," and the z-axis is "to the right," all determined by position and orientation during the calibration (see the Stimulus Recording section below). For the purposes of the data captured here, the WCS and BCS are identical, because the actor remained in a stationary position relative to the floor.

Each arm contained 16 bend/twist sensors, with the key sensor on each arm mounted on the outside of the actor's wrist. Through bend-and-twist information, the arm sensors measured both the position and orientation of the elbow and forearm. Using forward kinematics, translational data for the forearm were calculated relative to the orientation sensor, in an interface box placed on the upper arm near the shoulder. Each finger contained eight bend/twist sensors, for a total of 40 sensors, and was reported using forward kinematics relative to the wrist. The data concentrator converted the serial data output and transmitted the raw data through an Ethernet cable or wirelessly, through a wireless router, to a recording computer. The motion was viewed in real time on the recording computer and was recorded at the rate of 75 Hz.

Gestures and actions

A total of 119 gestures and actions were generated and categorized as either communicative or noncommunicative. Communicative gestures were defined as nonverbal behaviors that related to conveying or exchanging information with the recipient—for example, waving or giving a thumbs-up sign. These are commonly referred to as *emblems* (Ekman & Friesen, 1969; McNeill, 1985). Noncommunicative gestures were defined as object-oriented actions related to activities not intended to convey information to a recipient, such as mopping the floor or playing piano. These are commonly referred to as *pantomimes* (Ekman & Friesen, 1969; McNeill, 1985). This resulted in a total of 64 communicative gestures and 55 noncommunicative gestures.

Stimulus recording

A right-handed male without previous acting or sign language experience was selected to perform all of the actions.

Fig. 1 Front and back views of the wireless Measurand ShapeWrap III upper-body motion capture system, showing (**a**) a head orientation sensor, (**b**) a thoracic orientation sensor, (**c**) a pelvic orientation sensor, (**d**) two arm sensors, (**e**) four finger sensors for each hand, (**f**) a thumb sensor for each hand, and (**g**) the data concentrator

Movements were recorded using ShapeRecorder software version 4.06 (Measurand Inc., Fredericton, NB) on a PC computer running Windows XP (Microsoft, Redmond, WA). In their most raw state, the motion capture recordings were measurements of the sensors relative to the data concentrator (which was also attached to the actor's body). In order for these raw sensor data to accurately represent the position and movements of the actor's body parts, these data must necessarily be mapped onto a model of the actor in the recording software. The mapping of sensor data to the actor model during recording was achieved by a set of measurements prescribed by the manufacturer, which included measurements of various bones, distances between sensors and particular joints, and so forth. These measurements were entered into ShapeRecorder prior to motion capture. Prior to placing them on the actor, all sensors were calibrated according to the manufacturer's instructions.

Before recording, the actor was instructed to relax and to portray each action as naturally as possible. Each action was repeated several times; sometimes an action was later repeated additional times after other actions had been performed. The best version, in the opinions of the first and third authors, was used in the subsequent steps of stimulus production. A single action and its repetition were recorded in one take. At the start of every take, the equipment was calibrated to ensure a good correspondence between the sensors as represented in the real world and the sensors contained with the virtual model. During calibration, each axis of the BCS pointed in the same direction as the WCS. Since we had motion capture sensors only on the upper body, walking was disabled (i.e., the global motion component was removed), and the BCS and WCS were centered to the same point.

Each gesture in the communicative category started with a neutral pose: the actor standing with his arms to his sides, facing forward. The actor in this pose is shown in Fig. 2a; the representation of the same actor in ShapeRecorder is shown in Fig. 2b. Each communicative gesture also ended in the same pose. The noncommunicative actions could start and end in entirely different poses from this, depending on the natural flow of the individual gesture. During recording, objects were used to facilitate the creation of the noncommunicative stimuli. For example, while recording the noncommunicative gesture "drinking," the actor pretended to drink from an actual glass. The use of real objects made the stimuli appear more natural. Each movement was recorded a minimum of two times, with verbal feedback being provided by the researchers following each attempt. A video camera (HDR HC1, Sony Electronics) was set up 2 m in front of the actor and recorded the gestures at the same time as the motion capture. This was used for reference when importing the data into the animation-editing software, and for assistance during the editing stage to preserve the naturalistic quality of the stimuli.

The same actor recorded the communicative gestures on separate days from those on which the noncommunicative gestures were recorded. All recording parameters were kept constant across sessions; the body measurements taken during the first recording session were also used for the second session, to ensure the comparability of the two data sets.

When recording was over, the raw files were played back on ShapeRecorder to determine whether any missing or corrupted files were in need of recapture. Short clips of each action were exported offline from the ShapeRecorder software and saved in C3D format. C3D is a text file that stores data from the motion capture hardware at specific points on the actor's body. It can be imported into animation software, such as MotionBuilder 2009 (Autodesk, San Rafael, CA) for editing and rendering purposes. All of the editing and rendering were done on an Apple MacBook Pro laptop (Apple Computer Inc., Cupertino, CA) running Windows XP.

Postprocessing

For each gesture, a C3D file containing the action was imported into MotionBuilder 2009, and the markers were mapped onto an actor model. An actor model, shown in Fig. 2c, is an intermediate "skeleton" that serves as the source of motion within a subsequent character model. A character model is a 3-D object composed of a skinned model and the actor model skeleton. The character model can be animated in MotionBuilder once it is linked to a motion source through an actor model. This mapping consisted of assigning the markers in the C3D file to specific points on the actor model in MotionBuilder. This process necessarily altered the original motion capture data, since the actor model would not have the same physical proportions as the original human actor; however, each body/limb segment was scaled independently on the basis of the measurement data obtained from the human actor during recording (see above). For our study, we created a character, shown in Fig. 2d, with thirteen white spheres that marked the centers of the main joints, based on the procedure of Dekeyser et al. (2002). Twenty additional, smaller spheres were placed at the tip of each finger ($n = 10$), on each knuckle joining the finger to the hand ($n = 8$), and at the thumb joints ($n = 2$). Tartter and Fischer (1982) demonstrated that ASL signs presented using this set of point-light positions were readily understandable by native signers. Figure 3a shows the positions of the spheres on the hands and fingers. Our character model was created in Softimage 2008 (Autodesk, San Rafael, CA) and exported to a file format native in MotionBuilder.

Each gesture was then edited in MotionBuilder to correct misrepresentations in the joint positions and movement of the human actor, inaccuracies in the calibration of the equipment, or drift of the calibration settings over time. Further

Fig. 2 From top left, the panels in the figure illustrate (**a**) the human actor in the neutral pose, (**b**) the corresponding pose in the ShapeRecorder software, (**c**) the intermediate actor model created in MotionBuilder, and (**d**) the 33 spheres in the point-light model
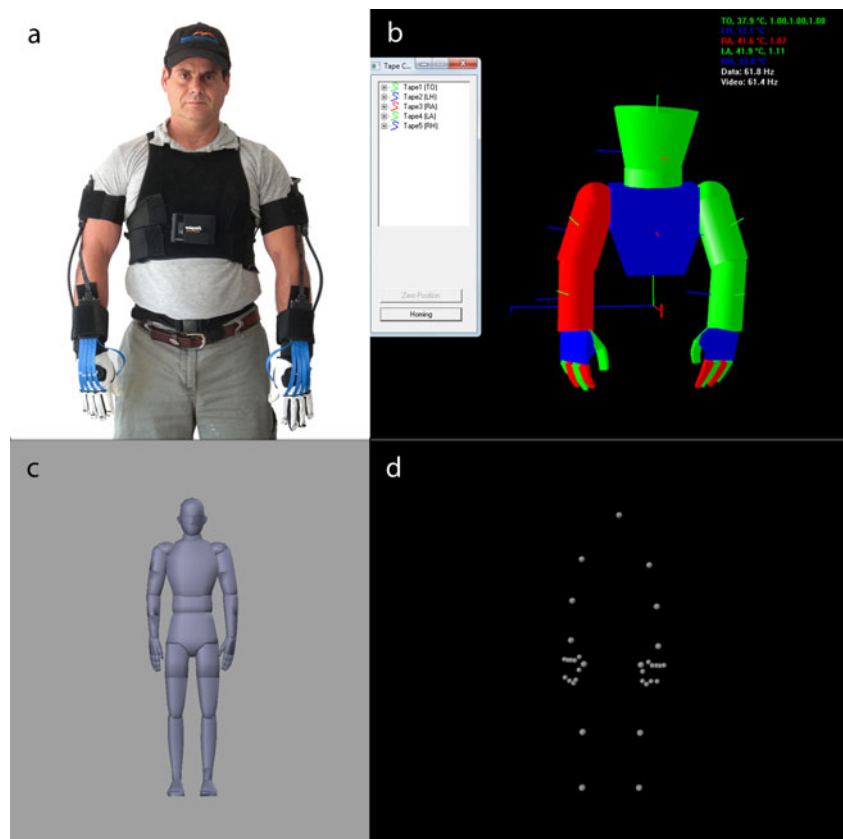
editing adjusted the movements of the model to make them more visually clear in the resulting animations. For example, an elbow position might be altered to prevent one arm from obscuring the hand of the other arm. The ensuing motion was then smoothed to avoid the appearance of "jumpy" dots. Throughout all movies, the dots representing the feet were "locked," making contact with the (virtual) floor. The knees were configured to maintain a natural pattern of motion, following the hips by a small fraction (10 %). An example movie is shown as Video 1 in the supplemental materials; a still frame from this movie is shown in Fig. 4.

Once processing was completed for a given gesture, the (x, y, z) coordinates of the 23 main joints (head, shoulders, elbows, wrists, finger tips, hips, knees, and ankles) were exported into a text file according to the TRC motion capture file format. The TRC files consisted of information
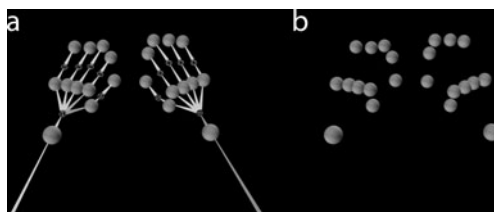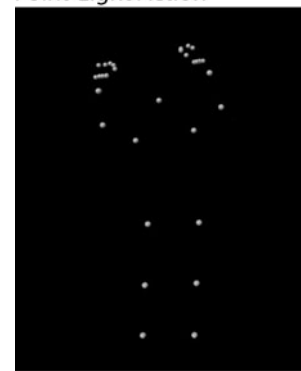

Fig. 3 (**a**) A close-up of the skeleton structure representing the actor's hands, showing the locations of the point-lights; (**b**) a close-up of the point-light model of the actor's hands, as used in the final stimuli

Fig. 4 Example frames from the point-light animations. The top panel shows the "Raise the roof" gesture; the bottom panel shows the scrambled version of the same gesture. The full movies are available as Videos 1 and 2, respectively

regarding the number of frames, the number of joints exported, and the coordinates of each marker over time. Compatible 3-D animation software packages are able to import data from this file format and to convert them into optical segments (the individual units whose movement is captured by the system, corresponding to actual bones of the human skeleton).

Scrambling

An attractive feature of point-light biological-motion stimuli is that they represent the coherent motion of the entire body through a limited set of dots. This allows us to test hypotheses concerning the perception of this coherent motion without influence from the appearance of the body, the actor's facial expressions, and so forth. It also allows us to scramble the starting positions of the individual point-lights in order to preserve the local motion vectors while disrupting the coherent percept of a moving human body (Grossman et al., 2000; Jokisch, Daum, Suchan, & Troje, 2005). This can be useful in testing hypotheses concerning global versus local motion, as well as in neuroimaging experiments in which one desires control stimuli that contain identical low-level visual information without the percept of biological motion (Grossman & Blake, 2002). With these considerations in mind, we produced a scrambled version of each of the communicative and noncommunicative gestures described above.

These scrambled videos were made as follows. In the first frame of each video, the y-coordinate of each point-light was inverted around its local x-axis, resulting in inversion of the figure. Then a new, random starting location was selected for the x- and y-axes (the z-axis remained unchanged). Having reassigned the starting positions of the point-lights, we subtracted the starting position from the original coordinate values of the point at each subsequent frame, thus preserving the trajectory of the point, but relative to the randomized starting position. As described above, each point maintained its original motion trajectory. Limits were set to ensure that each point remained viewable throughout the video. The inversion about the x-axis was performed because our subjective impression of movies produced by simply randomizing the starting positions of the point-lights was that too much of a percept of a human actor remained after scrambling without inversion. Inversion subjectively seemed to disrupt this percept sufficiently, an impression that was confirmed empirically (below). This may have occurred because we disrupted the perception of biological motion that might arise from local movements of individual dots relative to gravity (Troje & Westhoff, 2006) and/or because of the relatively large number of point-lights on the hands, as compared to other studies that have only randomized starting positions. An example frame from a

scrambled action is shown in Fig. 4. A new TRC file was then created from the new scrambled (x, y, z) coordinates and imported into MotionBuilder for rendering. The code to scramble each point was implemented in MATLAB 2007 (The MathWorks, Natick, MA).

Rendering

Once rendered, the point-light figure consisted of small white spheres on a black background, located approximately in the center of the visual field. The size of the spheres was chosen to allow visual separation of the individual spheres located on the hands at the desired output resolution of the animations. Spheres—3-D objects whose dimensions were indicated by shading—add depth cues that are not otherwise present with 2-D dots. These were used rather than simple dots because, in initial attempts at rendering, we found that the 3-D movements of the hands were not easily interpreted using 2-D points. Each nonscrambled gesture was rendered at one of four possible orientations. The model was facing either (1) toward the viewer (F), (2) 45º to the viewer's right (R), (3) 45º to the viewer's left (L), or (4) 45º to the viewer's right and tilted 10º downward, allowing for a view from slightly above the actor (A). The last view offered additional information from the point-light gesture in some cases, because of its downward perspective. The choice of each view was made on the basis of the animator's impressions of which view most clearly represented the gesture or action. Throughout the video, all 33 points were clearly evident to the viewer and were not masked by other points. Because view is meaningless when the coherent structure of the body is disrupted by scrambling, all scrambled videos were created from the forward-facing (0º) version of the action.

Stimulus formats

Movies The rendered actions and scrambled videos were rendered in the Apple QuickTime Movie (.mov) format, using the H264 codec, with a resolution of 640 × 480 pixels and a frame rate of 24 fps. The file sizes range from 50 to 150 kB, and each movie is between 1 and 4 s in length. These QuickTime files are provided as supplementary materials. Each video is categorized as either communicative or noncommunicative, and noncommunicative actions are further categorized as involving either the whole (upper) body or primarily the hands. The file names follow this naming convention: *action_category_view*, in which *action* is the name of the gesture (based on the most common name assigned by observers; see below); *category* is a letter, with C denoting a communicative gesture, B a whole-body noncommunicative gesture, and H a hands-only noncommunicative gesture; and *view* is F, R, L, or A (see above). The scrambled videos follow a slightly different naming

convention: scr_action_category, where "scr" denotes scrambled, and action and category are as described above for the nonscrambled videos. Since all scrambled videos were rendered from the same viewpoint (0 deg, or forward-facing), there was no need to code the view in the file names.

Text Text files are also provided in the supplement, containing the frame-by-frame coordinates of each point. These files are in the TRC format (tab-delimited text) and may be imported into any compatible animation-editing software or programming suite (e.g., MotionBuilder or MATLAB) for further modification or rendering. Further description of the TRC file format is provided in the README.txt file provided with the supplementary materials. These TRC files follow a similar naming convention to that used for the videos. However, since the view in each video is not essential, as the movement could be viewed from different angles once it was imported into the relevant software, the naming conventions are simply action_category for coherent biological motion, and scr_action_category for scrambled motion. Each text file is the same duration as the corresponding movie file, but with a frame rate of 30 fps. Each text file size ranges from 40 to 115 kB.

Normative data

Normative data were collected by asking a group of people naïve to point-light stimuli to identify the action depicted in each video. This was done to eliminate any gestures from the set that were difficult to identify.

Method

Participants A group of 20 undergraduate students (8 male, 12 female) participated in this experiment for course credit. The participants were naïve as to the purpose of the experiment and reported never having seen point-light animations before participating.

Stimuli and apparatus The stimuli presented to participants consisted of the 119 videos developed as described above. All videos were presented using DirectRT software (Empirisoft Corp., New York, NY) on a Mac Pro (Apple Inc., Cupertino, CA) computer running Windows XP (Microsoft Corp., Redmond, WA). The videos were presented on a 23-in. Apple Cinema Display LCD monitor (Apple Inc., Cupertino, CA) at a resolution of 1,920 × 1,280 and a viewing distance of 110 cm. Because of the high resolution of the screen, the videos were rendered at a resolution of 1,280 × 960 for this study, to ensure a good viewing size. Responses were polled using a standard USB Mac keyboard.

Procedure Participants were tested individually in a dimly lit room. Videos were presented in three blocks, each consisting of 39–40 randomly sampled stimuli. Each video was presented in the center of the computer screen, followed by a prompt to type a short description of what the participant thought the action was. The participants were instructed to type "don't know" if they could not recognize the action. Although participants viewed and described each movie at their own pace, they were not allowed to repeat any of the movies.

Scoring The scoring and analysis were based on the procedure used by Rossion and Pourtois (2004). Each movie was scored for the number of unique responses that it received, the most frequent response, and the proportion of participants who provided that response. The entropy statistic H (Shannon, 1948) was calculated to measure the agreement amongst the participants while controlling for the number of unique responses provided for each movie overall:

$$H = \sum_{i=1}^{k} p \cdot \log_2\left(\frac{1}{p_i}\right), \tag{1}$$

where k is the number of descriptions given to each video and p_i is the proportion of participants who responded with each name. Larger values of H represent more diversity (and less agreement) within the naming responses for a given movie, whereas perfect agreement (i.e., the same response provided by every participant) is represented by an H of 0.

The calculated H values were used to identify movies that were difficult to label or those that were strongly associated with more than one label. For this purpose, we rejected all movies with an H score greater than 1.5, as well as any for which "don't know" was the most common response. This resulted in the removal of 19 communicative gestures and 14 noncommunicative actions. The remaining 86 movies (43 communicative and 43 noncommunicative) constitute our stimulus bank. In the stimulus bank, the name of each movie file reflects the most common label given by our observers. The name of each communicative gesture is listed in Table 1, with a short description, a notation of whether or not the action is instrumental, and the H value. Table 2 includes the same data for noncommunicative, pantomimed actions.

Conclusion

We created a stimulus bank containing 43 communicative and 43 noncommunicative point-light actions suitable for

Table 1 Communicative, emblematic gestures included in the stimulus set

Action	Description	Instrumental	H	Length (s)
Air quotes	Raises both hands then bends index and middle fingers	–	0.29	2.00
"All done"	Wipes both hands against one another	–	0.47	2.36
Blowing a kiss	Kisses tip of hand then blows it	–	0.61	2.37
"Call me"	Fist to ear, extends thumb and little finger	–	1.19	2.20
"Calm down"	Both palms face down, and lifts both up and down quickly	–	0.00	2.36
Cheering fan	Both hands wave in the air	–	0.00	2.20
Clapping	Hands clap together	–	0.00	2.03
"Come here"	Hand sticking out, doing a pulling motion with fingers pointing up	–	0.29	2.36
"Come here for a hug"	Extends both hands and wags fingers inward before crossing the arms	–	0.99	2.06
Cross oneself	Tips of fingers to the head, then belly, then right followed by left shoulder	–	0.29	2.70
Disapproval	Leans back and crosses both arms	–	1.02	3.20
"Enough"	Palms together face down in front of stomach, then moved to the sides quickly	–	0.00	2.00
Finger wagging	Shakes the forefinger side to side	–	0.00	2.20
"Get out"	Points using index finger forward, then directs to the back using thumb	–	0.88	2.36
"Good idea"	Points and taps the temple	–	1.24	2.20
"Good job"	"Thumbs-up"—Arm stretches out and extends thumb	–	0.00	2.03
Hand at ear	Right hand cups right ear—"I can't hear you"	–	0.47	2.86
Hello	Right hand waves	–	0.00	1.70
"Hurry up"	Right palm faces inward and circulates outward	–	0.97	2.36
"I can't look"	Leans back with both arms tries to cover view	–	0.85	2.20
"I'm cold"	Shivers and hugs the body	–	0.00	2.20
"I'm not listening"	Both index fingers plug ears	–	1.32	2.03
"I'm sleepy"	Hands together next to the ear, then head tilts to the side	–	0.00	2.70
"I'm watching you"	Fixes the index and middle fingers on the eyes, then extends the hand and points the index finger outward	–	0.00	2.36
"It's hot in here"	Fans self using right hand	–	1.14	2.00
"Loser"	Extends the thumb and forefinger to resemble letter "L" on the forehead	–	0.85	2.70
Offer drink	Extends arm with fingers making a "C" shape	–	0.47	2.53
"One moment"	Index finger pointing upward, palm forward	–	1.52	2.70
"Over there"	Index finger points to the side	–	1.02	2.36
"Over your head"	Palm swoops over the head	–	1.46	2.03
"Raise the roof"	Both palms face up at head level and repeat an upward motion	–	0.00	2.86
"Rock on"	Thumb, index finger, and little finger stick out while fist shakes vigorously	–	1.02	2.03
Rubbing tummy	Hand rubs tummy in a circular motion	–	0.00	2.53
Salute	Performs military salute	–	0.00	1.86
"Shame, shame"	Wags finger back and forth	–	0.00	2.03
"Shhh"	Forefinger at mouth	–	0.47	2.03
Shooting	Extends thumb and index finger and pretends to shoot	–	0.00	2.23
Shrug	Both palms facing up at shoulder level, and shoulders raise slightly	–	0.29	2.36
"Smelly"	Right hand waves in front of nose and head tilts back slightly	–	0.00	2.53
Swooning	Right hand fans self and back of the left hand on the forehead	–	0.61	3.00
Thinking	Strokes the chin with thumb and forefinger	–	0.00	2.70
"Time out"	Hands form a "T" shape	–	0.29	2.53
Yawning	Covers the mouth briefly	–	0.00	1.86

"Instrumental" codes whether (+) or not (−) the action involves an object or objects. *H*, as described in the text, is a statistic indicating the degree of agreement amongst naïve raters. *H* = 0 indicates perfect agreement; higher values of *H* represent increasing heterogeneity of the descriptions provided by raters. Where used, quotes indicate an approximation of the verbal message intended by the action

Table 2 Noncommunicative actions included in the stimulus set

Action	Description	Instrumental	H	Length (s)
Bouncing a ball	Bounces ball twice	+	1.22	2.50
Boxing	Throws a punch then dodges	–	1.40	2.73
Brushing teeth	Brushes teeth back and forth	+	0.72	4.03
Buttoning a shirt	Buttons shirt upward	+	1.16	3.73
Call (phone)	Holds phone in hand, dials, and then brings to ear	+	1.46	3.20
Cleaning	Right hand moves vacuum or broom forward and back	+	1.19	3.77
Combing hair	Combs hair to the left side	+	1.39	3.03
Conducting orchestra	Moves hands as if directing an orchestra	–	0.00	3.20
Dancing	Random upper body dance moves	–	0.57	3.70
Drinking water	Brings cup to mouth, drinks, and puts cup down to chest level	+	0.72	2.87
Driving a car	Both hands hold the car steering wheel and steer car slightly	+	0.57	3.37
Drying one's body	Rubs the head and arms using a towel	+	0.29	4.03
Eating	Brings fork in right hand to mouth	+	0.47	3.73
Fishing	Casts fishing rod and reels back	+	0.00	3.67
Hammering a nail	Adjusts a nail and hammers it multiple times	+	1.34	2.37
Hanging clothes to dry	Flaps shirt, hangs on the clothesline, and pins both ends of shirt	+	1.15	3.77
Hold stomach	Hands hold stomach and body crunches forward	–	0.75	2.86
Juggling	Throwing and catching balls	+	1.44	2.10
Opening a bottle	Flaps a bottle cap upward	+	1.42	2.37
Opening a jar	Twists a jar cap	+	1.34	2.60
Paddling	Paddles a canoe	+	0.00	3.36
Patting a dog	Pats and strokes a dog	+	0.88	2.37
Petting a cat	Scratches the chin of a cat and strokes the back	+	1.00	3.07
Pick up a box	Bends down and lifts box	+	0.00	3.37
Picking up a paper	Picks up paper using thumb and index finger	+	0.47	2.53
Playing guitar	Left hand holds neck of instrument, right hand strums	+	0.00	3.03
Playing piano	Moves fingers of both hands on piano keyboard	+	0.93	3.03
Playing violin	Holds violin in left hand to shoulder and bow in right hand, then moves bow over violin	+	0.57	3.57
Pouring water	Pours water into a glass	+	0.57	3.37
Pull rope	Pulls rope toward self	+	0.00	3.70
Serving tennis	Throws ball upward and swings right arm	+	1.52	3.92
Sewing	Pokes needle into cloth and pulls needle upward	+	1.48	3.37
Shooting a basketball	Aims and shoots a ball into a hoop	+	0.00	2.87
Shoveling	Bends down to shovel snow and throws it overhead	+	0.00	3.87
Stirring	Stirs using the right hand, while the left hand holds the bowl in place	+	0.85	2.20
Sweeping the floor	Both hands hold a broomstick and sweeps across	+	0.29	2.40
Swimming	Freestyle swimming strokes	–	0.00	3.20
Throw	Aims and throws spear forward	+	1.46	3.33
Throwing a snowball	Makes a snowball and throws it	+	0.72	2.53
Washing	Scrubs a wall in a circular motion using a sponge	+	0.99	2.37
Washing hands	Soaps and washes the hands	–	0.72	2.03
Writing	Writes using the right hand	+	1.29	3.37
Yell	Both palms cup the mouth	–	0.92	2.36

The details are as for Table 1

use in behavioral and neuroimaging research. We have demonstrated that all actions selected for inclusion in this stimulus set are highly recognizable; in many cases, the desired label was the sole response provided by all

participants. The points of light represented on each finger are unique among the point-light stimulus sets that are readily available, making these stimuli well-suited for investigation of communicative, emblematic gestures. As the noncommunicative pantomimed actions share similarities to previously released stimulus banks (e.g., Vanrie & Verfaillie, 2004), this set of stimuli is suitable for a wide range of studies, including those designed to investigate the role of the hands in pantomimed actions and to compare communicative and noncommunicative hand actions.

References

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics, 23,* 145–152.

Cutting, J., & Kozlowski, L. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9,* 353–356.

Dekeyser, M., Verfaillie, K., & Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behavior Research Methods, Instruments, & Computers, 34,* 375–382. doi:10.3758/BF03195465

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

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1,* 49–98.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192. doi:10.1038/nrn1057

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

Grossman, E., 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,* 711–720. doi:10.1162/089892900562417

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

Jokisch, D., Daum, I., Suchan, B., & Troje, N. F. (2005). Structural encoding and recognition of biological motion: Evidence from event-related potentials and source analysis. *Behavioural Brain Research, 157,* 195–204. doi:10.1016/j.bbr.2004.06.025

Ma, Y., Paterson, H. M., & Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods, 38,* 134–141. doi:10.3758/BF03192758

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350–371. doi:10.1037/0033-295X.92.3.350

Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., & Neville, H. J. (2002). A critical period for right hemisphere recruitment in American Sign Language. *Nature Neuroscience, 5,* 76–80.

Newman, A. J., Newport, E., Supalla, T., & Bavelier, D. (2007). Neural systems involved in the comprehension of American Sign Language (ASL) and non-linguistic gesture: An fMRI study. *Journal of Cognitive Neuroscience, 19*(Suppl.), 288.

Newman, A. J., Supalla, T., Hauser, P., Newport, E., & Bavelier, D. (2010a). Dissociating neural subsystems for grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences, 107,* 7539.

Newman, A. J., Supalla, T., Hauser, P. C., Newport, E. L., & Bavelier, D. (2010b). Prosodic and narrative processing in American Sign Language: An fMRI study. *NeuroImage, 52,* 669–676.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33,* 217–236. doi:10.1068/p5117

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, 17,* 379–423.

Tartter, V. C., & Fischer, S. D. (1982). Perceiving minimal distinctions in ASL under normal and point-light display conditions. *Perception & Psychophysics, 32,* 327–334.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology, 16,* 821–824. doi:10.1016/j.cub.2006.03.022

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers, 36,* 625–629. doi:10.3758/BF03206542

White, N. C., Fawcett, J., & Newman, A. J. (2009). Biological motion perception: An ERP study on the functional role of the N300 component. *NeuroImage, 47*(Suppl. 1), S156.