A technique for continuous measurement of body movement from video

Caitlin Hilliard[1] · Susan Wagner Cook[1]

Abstract The movements that we make with our body vary continuously along multiple dimensions. However, many of the tools and techniques presently used for coding and analyzing hand gestures and other body movements yield categorical outcome variables. Focusing on categorical variables as the primary quantitative outcomes may mislead researchers or distort conclusions. Moreover, categorical systems may fail to capture the richness present in movement. Variations in body movement may be informative in multiple dimensions. For example, a single hand gesture has a unique size, height of production, trajectory, speed, and handshape. Slight variations in any of these features may alter how both the speaker and the listener are affected by gesture. In this paper, we describe a new method for measuring and visualizing the physical trajectory of movement using video. This method is generally accessible, requiring only video data and freely available computer software. This method allows researchers to examine features of hand gestures, body movement, and other motion, including size, height, curvature, and speed. We offer a detailed account of how to implement this approach, and we also offer some guidelines for situations where this approach may be fruitful in revealing how the body expresses information. Finally, we provide data from a small study on how speakers alter their hand gestures in response to different characteristics of a stimulus to demonstrate the utility of analyzing continuous dimensions of motion. By creating shared methods, we

hope to facilitate communication between researchers from varying methodological traditions.

The human body affords a multitude of movements, and many of these convey meaning or information. For example, people gesture when they talk, nod their head when they agree and shake their head when they do not, sit up straight when they are attentive, and bounce their leg when they are bored or nervous. Although it is often the presence of a particular behavior that is of interest, movements can vary in a continuous fashion in ways that might also be informative. Hand gestures may be raised higher when the information is new to the listener (Hilliard & Cook, 2015), a nod may be larger when the listener vehemently agrees with something, and leg bouncing may become slower as the speaker becomes less anxious. Movements of the body are of interest to researchers across psychology, linguistics, anthropology, dance, communication, and other fields. Despite this robust interest in the movements of the human body, many systems that code or measure movement group movements into categories and thus may not capture the richness present in movements. For example, gesture researchers have focused on categorical coding systems without developing ways of measuring continuous variation, even though theoretical approaches in gesture research have emphasized the potential of gesture for analog and iconic representation that would depend on continuous variation (Hostetter & Alibali, 2008; McNeill, 1992; De Ruiter, 2000). Here, we motivate and describe a method for coding body movements from video that is applicable to coding a wide range of movements along continuous dimensions, and we

✉ Caitlin Hilliard
caitlin-hilliard@uiowa.edu

[1] Department of Psychological and Brain Sciences, University of Iowa and Delta Center, E11 Seashore Hall, Iowa City, IA 52242, USA

illustrate the utility of this approach for gesture research in a small experiment.

The need for capturing movement continuously

Speech-accompanying gestures as one example for body movement

Hand gestures are extremely prevalent during language production. When people speak, they also gesture with their hands. Hand gestures are typically produced rhythmically along with the speech that they accompany and they are related in meaning to the accompanying speech (Kendon, 2004; McNeill, 1992). Gestures are produced in precise temporal coordination with speech, with the stroke of the gesture typically slightly preceding and overlapping the word or phrase that it represents or emphasizes (McNeill, 1992a, b). Speech and gesture can be considered an integrated system for communication, from both a speaker and a listener perspective (Kelly, Ozyürek, & Maris, 2010; Kendon, 2004; McNeill, 1992).

Despite the tight coupling of gesture with speech, the way that spontaneous gesture communicates information via hand movements is different from spoken language. While spoken language largely encodes a message sequentially, with a series of phonemes leading to words produced over time, gesture can express multiple aspects of a message simultaneously. For example, consider the case of a speaker describing a child going down a slide. In speech, this might manifest as a series of discrete segments ordered in a syntactically meaningful way: "The child went down the slide." The accompanying gesture could be one finger moving downward in a spiral motion to depict the child sliding down a tornado slide. This gesture illustrates the message in a single fluid form produced in conjunction with multiple words from the spoken message. Although the position of the hands varies with time, other features do not. Instead, the gesture simultaneously depicts multiple features of the intended message. The motion of the hand over time represents a single downward trajectory. The shape of the finger may encode some information about the child's body position when sliding. Moreover, the gesture expresses information that was not contained in the concurrent speech; that is, that the slide the child used was spiral rather than straight down and perhaps even the speed at which the child moved down the spiral. All of this information is simultaneously contained in a single movement of the hand.

Importantly, slight variation in any of the dimensions would likely alter the meaning expressed in gesture (Beattie & Shovelton, 1999). If the finger had been slightly bent, this might depict that the child was sitting rather than lying down. Similarly, changing the steepness of the downward slope or the degree to which the gesture is spiraled could lead to

different interpretations of the event and the nature of the movement. It is known that observers are sensitive to subtle variations in movement. For example, the weight of a box lifted by another person can be perceived via the visual motion information (Runeson & Frykholm, 1983), the expertise level of musicians can be detected from their body movements (Rodger, Craig, & O'Modhrain, 2012), and numerical information in the environment is expressed in the scaling of one's grip to objects during reaching movements (Chiou, Wu, Tzeng, Hung, & Chang, 2012).

Although research on movement dynamics has typically used motion-tracking systems to obtain detailed information, researchers in other traditions have often not had access to such systems, or have been unable to implement them due to the nature of the research questions. For example, current measurement techniques in gesture research typically capture continuous variation in gesture form in a categorical fashion. In the field of hand gesture, conventions from sign language research have informed the methods used to capture movement properties, leading to categorical coding. Sign language signs are often annotated using Stokoe, Casterline, and Croneberg's (1976) system. In this system, three features of the hand are annotated: handshape, location, and motion. There is also a fourth, later-added feature, orientation, that is less frequently analyzed in gesture research. Gesture coding techniques typically assign gestures to categories on the basis of these dimensions; each gesture is coded as having a particular motion, orientation, handshape, and location, as well as other categorically coded aspects like handedness. Each gesture is typically assigned to one category for each dimension (e.g., for handedness: right, left, both; for motion: lateral, vertical, etc.). For example, in the McNeillian system, gesture phases are identified, and then each phase is coded by following a strict descriptive system of these phases (e.g., Kita, Van Gijn, & Van der Hulst, 1998). This provides a description of the gesture form; from the annotation for each gesture, the form of the original gesture can be approximated and recreated. This technique has been very useful for describing gesture in a variety of research approaches, including in qualitative analyses (e.g., Cardona, 2008) and in explorative situations (e.g., Alibali, Evans, Hostetter, Ryan, & Mainela-Arnold, 2009). This system along with a typology of gesture (McNeill, 1992) has been used to document wide variation in gesture production across speakers, topics, and contexts.

Other research traditions examining human movement have also developed coding schemes that are categorical. The Bernese system (Frey, Hirsbrunner, Florin, Daw, & Crawford, 1983), for example, describes the positioning of the limbs along three Cartesian axes over time and thus details nearly every potential degree of freedom in movement. Similarly, Laban Movement Analysis describes movements according to categories encoding dimensions of body, movement, shape, and space (Zhao & Badler, 2001). The Body

Action and Posture Coding System (BAP) is also executed similarly, but captures and categorizes the function of movements along with the articulator used and form of movement produced (Dael, Mortillaro, & Scherer, 2012). These annotation systems allow for visualization, reconstruction, and interpretation of the annotated movements. These systems and the gesture measurement techniques described above provide multiple pieces of information about each movement and do so by describing movements according to categorical features.

Continuous coding schemes should be applied to continuous dimensions of movement

Of course, the way we transcribe our data necessarily and implicitly reflects our interests and hypotheses and constrains the inferences that can be made (Ochs, 1979). Despite the utility of the aforementioned coding schemes, they fall short when they are applied to the properties of the movement themselves. Describing motion features categorically does not allow one to capture continuous variation potentially available within all the aforementioned dimensions. Motion can clearly vary within a category; for example, describing a movement as "lateral" says nothing about the directionality, the distance traveled, or the degree of angle from horizontal. Even handedness, which is seemingly necessarily categorical, can vary continuously; production of a large movement with the dominant hand can be accompanied by small but similar movements on the opposing hand. Thus, across coding systems, many movements that have identical annotations in all categories will have been variable in form in their original production.

Indeed, even measurement techniques that explicitly intend to capture continuous dimensions of hand movements are often fundamentally categorical, perhaps in part because of the tradition and familiarity of categorical coding systems. For example, McNeill (1992, pp. 86–89) described a method for capturing the position of hand gestures in gesture space in which the gesture space is divided into several discrete segments, and gestures are categorized as falling into a particular segment (e.g., Kuhlen, Galati, & Brennan, 2012), some of which are considered central and some peripheral. Similarly, Goldin-Meadow, Mylander, and Franklin (2007) describe a handshape coding system in which the distance between the thumb and the fingers is categorized as touching, small, medium, or large. However, given the great deal of individual differences in the natural gesture space in which each person gestures (Priesters & Mittelberg, 2013) and the potential of the thumb and fingers to assume any one of an infinite number of distances, analyzing across these sort of categories necessarily disregards potential richness and variability in the shape and form of gestures.

Thus, despite a theoretical emphasis in the field of gesture studies on analog and continuous aspects of the hands, the ability to investigate continuous movement properties of gesture has been limited by the lack of measurement techniques that can do so. Rather than treating body movements – and their movement properties – as though their production is categorical, quantitative analysis can enable us to consider continuous variation in these movements.

Using non-continuous coding necessitates categorical analysis

Aside from the theoretical implications of coding continuous variation in body movement categorically, using categorical coding systems also constrains the type of analyses that can be conducted effectively. Although it is common practice to transform categorical data into continuous variables for analysis, this can introduce spurious results and decrease power (Jaeger, 2008). It is better practice to use statistical models that capture the underlying structure of the data, either categorical or continuous. Continuous data has some advantages for analysis. Statistical methods for continuous data are well developed and take advantage of the finer detail and greater information inherent in continuous measurement, allowing for stronger inferences with fewer data points (Zhao & Kolonel, 1992). Given that coding movement is often time-intensive, it seems prudent to attempt to maximize the information gained for analysis.

Continuous coding is not conducted due to constraints in measurement techniques

Why do we continue to analyze continuous properties of movement categorically? One of the greatest difficulties with capturing continuous information in body movement is the laborious nature of doing so. In the field of gesture studies, one of the first analyses capturing motion was conducted by Efron (1972). In his seminal examination of cultural differences in gesture production, Efron detailed the motion and size of gestures produced by individuals from four different ethnic groups using a series of still images and some motion pictures. To document his findings, an artist produced elaborate drawings of the hands with small dots detailing the trajectory of each hand. Despite being incredibly thorough, this intensive style of coding is likely too laborious and time-consuming for contemporary analysis of gesture.

Quantitative descriptions can now be acquired more readily with motion-tracking technology (Priesters, 2013). This has obvious advantages: it captures a comprehensive record of movement over time, is does not require extensive coding after collection, and can have a high level of precision. However, motion tracking often requires that participants wear sensors on their appendages to track all movement produced and this may very likely clue participants in to the goals of the study,

which is often not acceptable in gesture research and cannot be done when researchers are interested in examining live performances or other actions in context. Newer motion-tracking systems make the use of sensors unnecessary by using multiple cameras that integrate videos to create a three-dimensional motion signal. This can be extremely useful in laboratory environments. However, these systems may be of less utility in naturalistic situations outside of the laboratory; the logistics of setting up a motion-tracking system outside of the laboratory (e.g., in a classroom) are complicated and may have issues capturing the movement of multiple people at various locations and orientations within a space. Additionally, motion tracking can obviously not be used post-hoc: if data have already been collected, motion-tracking systems are of little use. Finally, motion-tracking systems are often expensive and generate extremely dense data sets that further complicate analysis. Thus, even if researchers are interested in capturing continuous aspects of movement with automated systems, there are considerable barriers to doing so.

In order to facilitate continuous coding of movement, an accessible and simple method to do so is required. In the rest of this paper, we describe a method that we believe satisfies this need by allowing the researcher to capture a motion signal of a multitude of articulators – or parts of the body that can produce movements - using technology and software that is readily available. Although there are a variety of steps in our process, it is relatively quick and efficient. Advances in image processing will likely streamline the process even further.

Measuring body movement continuously: A new approach

We have developed an approach that quantifies body movement trajectories in a low-cost but fine-grained fashion. Our measurement technique provides a two-dimensional motion signal of each movement of interest, captured frame-by-frame. When this method is applied to gesture production data, the position of each hand in each frame of video is annotated by simply clicking on a stable point of the hand in each image and recording the screen position of the click. This outputs an x-y coordinate for each articulator: the positioning of the articulator in each frame. Thus, this signal provides information about changes occurring within each individual gesture over time. From this signal, size, position relative to the body, speed, curvature, and trajectory can be analyzed, both within and across different categories of movements. Categories like handshape can be processed in a similar way, depending

on the dimensions of interest, and combined with the continuous motion information to provide a multi-dimensional record of the hand configuration and position over time. Thus, our approach builds on and extends prior work using categorical coding schemes. By adding continuous measures, we can explore characteristics of gestures and other body movements in greater detail.

In order for this approach to be used, a video recording of the movement of interest is needed, something that is already a necessity for studying body movement. It is preferable for videos to have a relatively stable camera position across all videos considered in an analysis, since the quantitative output is coordinate points on a computer screen. If the camera angle varies, this introduces variability into the motion signal across camera positions, which may be a problem depending on the question of interest. Although it is possible to correct for deviations in camera position by transforming the data points to a normed space for all participants or by including reference objects in the video image, this requires additional steps and is likely to introduce additional noise into the data.

Annotation of body movement trajectories over time is a multistep process that requires that the movements of interest be identified and the positioning and/or shape of the articulator(s) annotated on each frame of video. The process we use to carry this out is described in detail below, although of course this approach can be implemented in a variety of ways.

The measurement process

Identifying gesture timing information Although one could annotate an entire stream of video and identify movements of interest strictly from their motoric properties, it is often desirable to first identify when movements of interest occur in the video data according to predetermined criteria. This will considerably reduce the amount of information to be annotated – especially since researchers in many fields already have well-established criteria for identifying behaviors of interest. For our coding of gesture production data, we first identify gestures using ELAN (EUDICO Linguistic Annotator) video annotation software (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) and we then use the timing information from ELAN to extract video frames from only the gestures of interest for subsequent coding (Fig. 1, Step 1). After the movements of interest have been identified and categorized, these movements are further processed so that the continuous trajectory information can be coded efficiently.
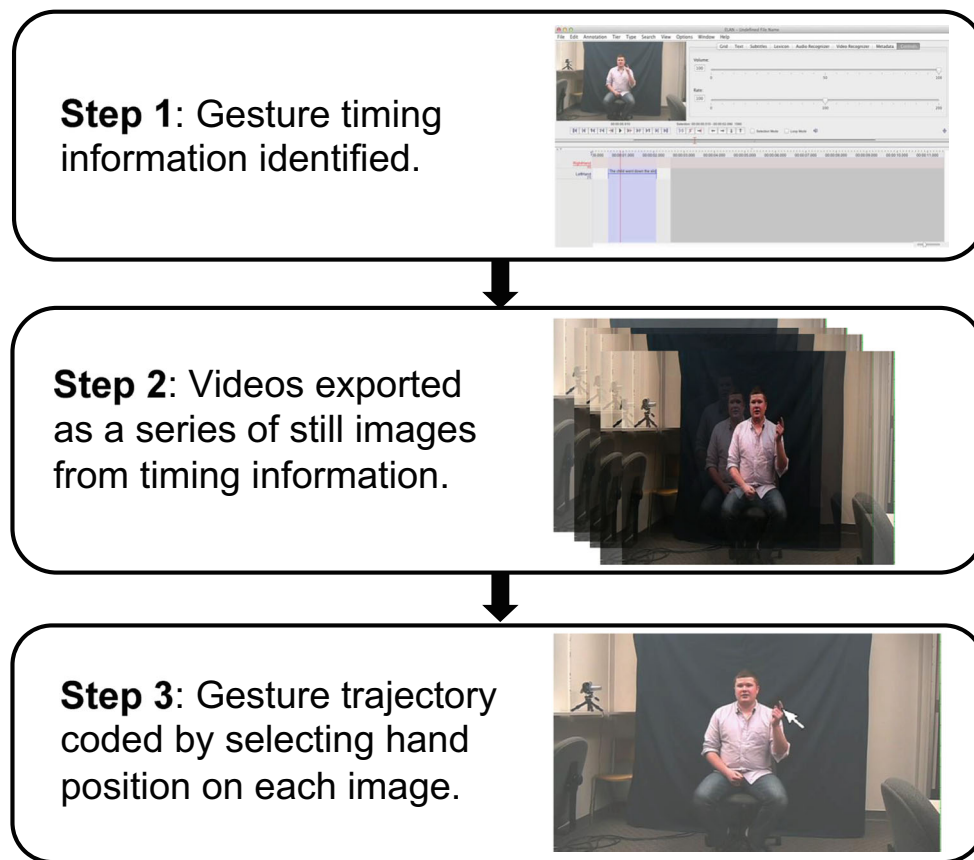
Fig. 1 Schematic depiction of the coding method

Exporting videos as series of images These annotated movements are processed into images for analysis. To reduce the amount of information and save time, we have found it helpful to down-sample the video when we process it into a set of images (Fig. 1, Step 2). We have been sampling videos at the rate of ten frames per second, although alternative sampling rates may be desired depending on the level of detail appropriate to one's hypothesis. Thus our process identifies movements of interest in the original video and then each individual movement is transformed into a series of still images. This information can be quickly obtained from the images using any program that presents the images as stimuli and gathers mouse or keyboard responses, automatically saving the coded data. For example, to capture trajectory, the x and y coordinates of a mouse click are recorded for each image, yielding information about how each hand is moving over time (Fig. 1, Step 3). Of course, multiple articulators (e.g., hand and head) can be annotated for each movement for later comparison and analysis.

Coding movement trajectory The result of this process is a series of data points detailing how each articulator moves in two-dimensional image space during each individual movement along with any categorical information also annotated for each frame (Fig. 2). Prior to analysis, it is often useful to invert the y-coordinates; on computer monitors, the origin point is typically in the top left corner, and thus lower coordinates actually denote higher positioning on the screen. This can be corrected by simply subtracting the raw y-coordinates from the overall height of the monitor in pixels. It is also possible to account for variation in participant height by choosing norming points on each body and transforming all the coordinates collected into a common space.

Analyzing multiple articulators

It is often the case that multiple parts of the body move simultaneously. With this measurement method, the motion signals from multiple articulators can be easily integrated and compared. One obvious example is the temporal coordination of hand gesture and speech. Using ELAN (Wittenburg et al., 2006), one can integrate information obtained via this technique with analysis of speech from Praat (Boersma & Weenink, 2015),
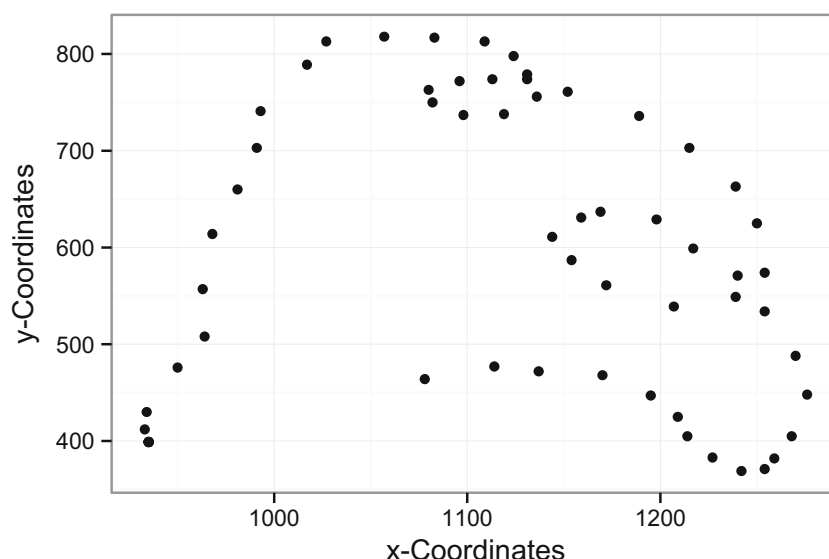
Fig. 2 A plot showing the spiraling trajectory of a gesture accompanying the sentence "The child slid down the slide"

allowing one to capture synchrony at the level of the frame. By making multiple passes through the images, researchers can easily investigate multiple articulators.

With multiple passes through the data, this method could also be readily extended to capture handshape. Although handshape has typically been coded according to categories derived from American Sign Language (ASL; following Stokoe et al., 1976), it is not clear that gesture handshapes are perceived categorically. In fact, it is even possible that continuous dimensions may underlie handshape perception in ASL (Stungis, 1981). By annotating the position of multiple joints or points on the hand, researchers can capture the shape and size of the hand in a more continuous fashion, allowing investigation of the possibility of continuous coding of information in handshape, as well as the possibility of change in handshapes within a single gesture.

Incorporating categorical data during movement measurement

As alluded to earlier, categorical information will often also be of interest along with continuous data. Conveniently, traditional categorical coding can also be applied with this technique, by coding each frame categorically rather than continuously. This can be done efficiently by recording a keystroke for each frame rather than a mouse click. The advantage of this method of collection is that categorical information can now be considered in conjunction with time and with other continuous measures like trajectory. For example, proportions of time that the hand is in a particular category

can be analyzed for a single gesture (i.e., a movement might be 60 % C handshape and 40 % V handshape).

Below we provide examples of some dimensions of movement that our method can capture. To generate these examples, we conducted a small study using stimuli predicted to elicit a variety of specific form-based alterations in gesture production.

Capturing continuous variation in gesture form: An example

As mentioned earlier, hand gesture is particularly well suited to communicate multiple properties simultaneously. To date, many of the studies examining how gesture communicates this type of information have focused on single properties of gesture. However, it is very likely that multiple properties of a gesture are expressing relevant information simultaneously (Senghas, Kita & Özyürek, 2004). Returning to our earlier example of the child going down the slide, the spiraling gesture communicated multiple features of the event. To demonstrate the functionality of our coding system, we examined how speakers modulate multiple components of their gesture production when describing the movement of a stimulus. Our study followed Shintel, Nusbaum, and Okrent, (2006), who examined the use of analog representation in the vocal channel. In their study, participants spoke about a dot moving to the left or right at varying speeds, saying, "It's going right," or "It's going left." The speed of the dot was irrelevant to the participants' task; nonetheless, Shintel et al. (2006) found that participants encoded speed, speaking more quickly for

dots that moved more quickly. Moreover, listeners were sensitive to this information.

Our study asked whether participants similarly encoded speed information reliably in gesture. We also asked whether participants spontaneously encoded additional irrelevant information in gesture, specifically the location of the dot in space, given prior work suggesting that gesture is effective for communicating spatial information. Although it has been posited that gesture reflects the properties of the object or action that it represents (Hostetter & Alibali, 2008; McNeill, 1992, to date there are no empirical studies that we know of that have directly investigated this issue for these dimensions.

Methods

Four right-handed student participants at the University of Iowa viewed 24 short videos of a dot moving left or right on the computer screen, and their task was to describe which way the dot moved (left or right) in a full sentence. Participants were asked to gesture on each trial. They could not begin speaking until the dot finished moving, to prevent syncing of speech, gesture, and dot movement.

In addition to varying the direction it moved (left or right), which was relevant for the participants' communicative task, the dot also varied in the speed (four possible speeds) and the global position on the screen (at the top, middle, or bottom of the computer screen). Position and speed were not relevant for the communicative task. On each trial, the dot was visible for 2 s on the horizontal midpoint of the screen at the top, middle, or bottom before it began moving. The dot always traveled the same distance (600 pixels), but did so in 1, 2, 3, or 4 s. Accordingly, the depicted speeds were 600 pixels(p)/s, 300 p/s, 200 p/s, or 150 p/s. Thus, there were multiple properties of the stimulus that could be encoded in gesture.

Prior to analysis we identified the stroke of each gesture (McNeill, 1992, which was the only phase that we analyzed here. We did not code the preparation and recovery phases (Kendon, 2004), as we did not predict these phases would be modulated along the dimensions of interest.

Analysis We analyzed the movement data with linear mixed effect models predicting the feature of interest with fixed effects for the relevant features of interest. The random effect structure was determined by log-likelihood ratio testing to determine the maximal random effect structure justified by the data. Below we report coefficients and t-values in the absence of p-values, as there is no consensus regarding how degrees of freedom are calculated for mixed regression models (Bates, Maechler, Bolker, & Walker, 2015). We assume that all t-values with an absolute value above 2 correspond to significant findings.

Results

Height/positioning This details where in the gesture space each gesture is produced, relative to the body. To calculate this, we first had to locate a stable point on the body so we could norm all trajectories to this point to account for variation in camera angle and body size. Given that all of our participants were sitting, we chose the groin area (while participants often move their head and trunk while communicating, the area of their body that is in direct contact with the chair tends to remain stable). These norming points were collected in the same way as described above. The x and y coordinates of the norming point for each participant was subtracted from the x and y coordinates of all of the gestures for that participant.

We predicted that the higher the dot was depicted on the screen, the higher that the gesture describing it was produced relative to the participant's body. Our model predicting height had a fixed effect for the height of the object described on the screen and a random effect for participant. We found that gestures produced for dots located at the top of the screen were reliably higher than those in the middle of the screen (β = 46.3, t = 3.87) and those in the bottom of the screen were reliably lower than those in the middle (β = -686, t = −5.69; Fig. 3). When including trial number as a fixed effect in the model, we also found a significant lowering of gestures across the 24 trials regardless of the height of the dot in the video (β = −1.76, t = −2.54; Fig. 4). Thus, participants' gesture heights were not only semi-veridical representations of the motion of the dot, but were also a function of how far into the experiment they were.

This finding makes the analyses possible with our coding scheme particularly evident. Had we coded gesture height with McNeill's (1992) gesture space description, we likely would have not detected the general lowering effect; while there was lowering for each of the heights present in the stimulus, gestures very often fell within the same descriptive area in McNeill's description (Fig. 3). Thus, the ability to capture small variation in gesture form was necessary to detect this effect.

Not surprisingly, we also found an effect of the placement of gestures horizontally as a function of the direction of
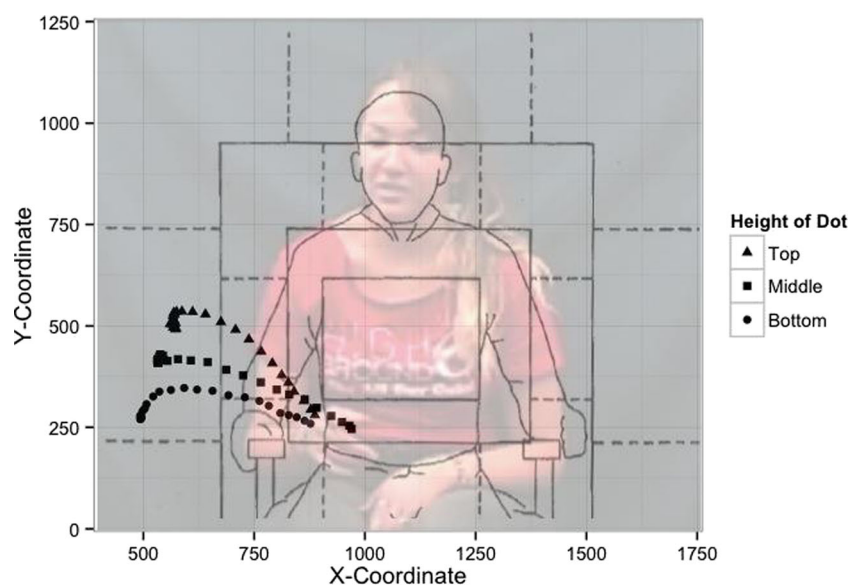
Fig. 3 The trajectories of three gestures produced by one participant describing the same movement (in terms direction and speed) but with the stimulus dot located at the three possible heights. The McNeill (1992a, b) gesture space grid is also depicted to illustrate that all three gestures, while reliably different in height, are produced in the same general gesture space

motion that the gesture was depicting. Using a model that had a fixed effect for direction and random effect for participant, we found that gestures were significantly more left relative to the body when produced for dots moving to the left (β = 379.18, t = 39.2).

Speed We also predicted that participants would adjust the speed of their gestures to reflect the speed of motion observed. We calculated the speed of each gesture by dividing the distance traveled during a gesture by the time it took to complete the gesture. Of course, the distance between each of the points of the gesture was calculated, as calculating from the end points disregards any deviation from a straight trajectory.

Again, our prediction was confirmed; a mixed effect model predicting velocity (p/s) with a fixed effect of speed of the dot and a random effect of participant revealed that the less time it took the dot to move across the screen, the faster the hand moved (β = −.012, t = −2.71; Fig. 5). Beyond the use of motion tracking systems, this is the first method of coding that we know of for capturing gesture speed.
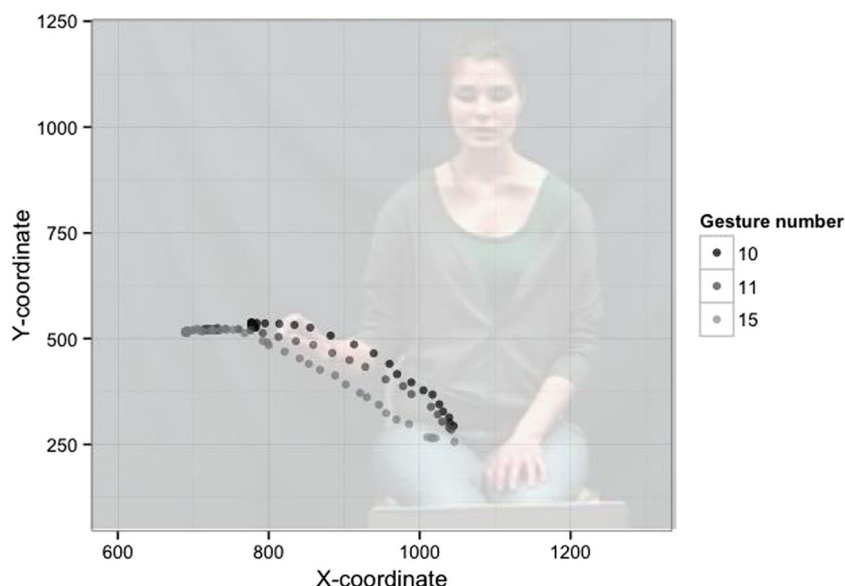

Fig. 4 The trajectories of three gestures produced by one participant describing dot movement at the same height and direction, but across different trials. Only the speed of the stimulus varied in each of the trials. Despite describing the same movement at the same location on the computer monitor, the gestures are gradually lowered in the gesture space across descriptions
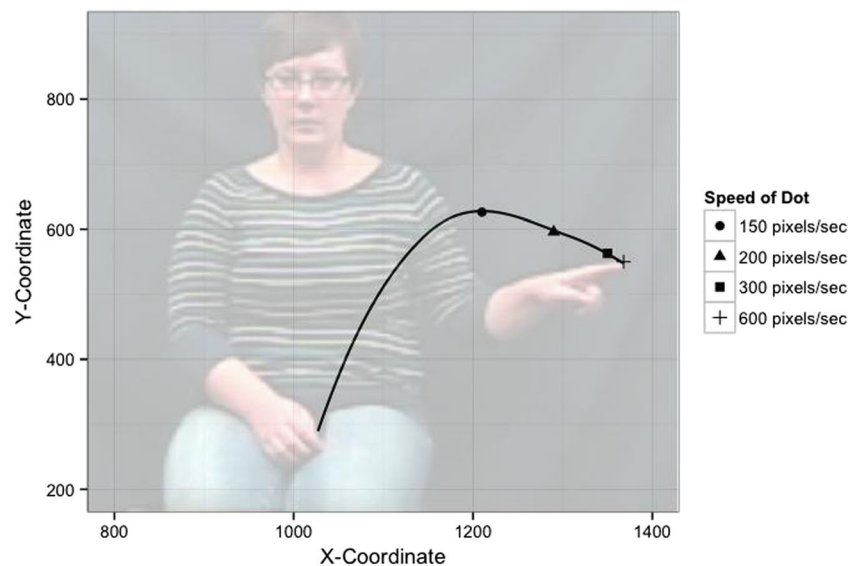
Fig. 5 Speed of movement depicted relative to a single example trajectory for a dot moving at the fastest speed. Each symbol represents the distance that would have been traveled by the hand if it were moving at the average speed produced by the speaker when depicting dots moving at slower speeds

Size The size of gestures was determined by calculating how far the hand moved in pixels during each gesture. We analyzed only the size of the gestures horizontally, as the stimuli depicted horizontal motion. We determined the maximum displacement during each stroke by subtracting the minimum x coordinate (extracted from our coding system) from the maximum x coordinate for each gesture (also extracted from our coding system). In cases of non-linear movements, the distance between each point in a movement can be calculated and summed to determine total size. Curves can also be fit to the data when appropriate, and characteristics of these curves can be studied (see Cook & Tanenhaus, 2009; Hilliard & Cook, 2015). Note that this approach to calculating gesture size allows the placement of the hand relative to the body to be unconfounded from the location of the gesture in space. In prior work, gesture size has been determined by how much the hand deviates from the center of the gesture space

(identified as the Center-Center in McNeill's (1992, pp. 89) depiction of gesture space). A series of rectangles growing in size is drawn around the central area of the gesture space and each gesture's size is determined by choosing which rectangle it reached (see Fig. 3). However, this makes directly comparing gestures that are produced in different areas of the gesture space difficult – and we have already established that leftward and rightward gestures are produced in different regions of space. By continuously capturing the motion signal, we can determine that two movements produced in vastly different areas of the gesture space are the same size.

We analyzed size in a model predicting lateral displacement in coordinates with fixed effects of direction, height, and speed and a random effect of participant. We had no explicit predictions about size, as we did not vary the distance traveled by the dot in any videos. Surprisingly though, we did find an effect of size as a function of direction; gestures
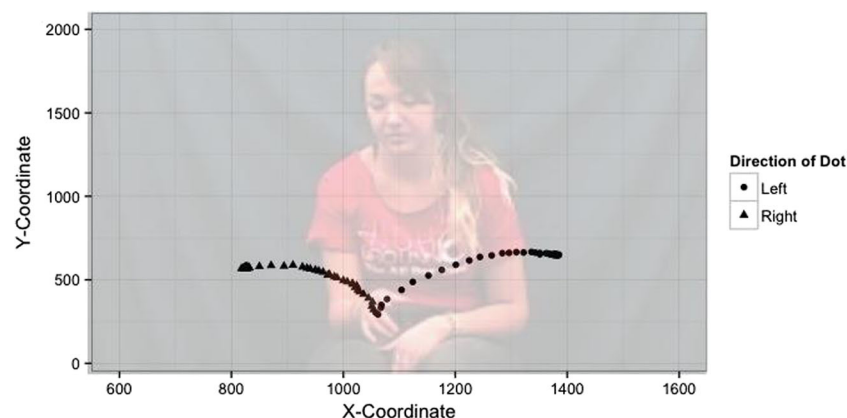

Fig. 6 Two trajectories for one participant's description of dot movement that varied only in the direction (left vs. right) that it moved. Note that the hand travels a greater distance for the leftward trajectory

moving leftward were larger than those moving rightward (β = 127.98, t = 23.27; Fig. 6). Since all participants were right-handed, they had to move their hand more distance overall when describing leftward movement to reach the same relative location on that side. One interpretation of this novel finding is that speakers are depicting distance as a function of the end point relative to the body, rather than the actual distance travelled.

Although we have focused on just the three main variables of interest in the present experiment, many additional continuous variables can be extracted from the motion signal and analyzed. For example, the curvature, duration, resting position between gestures, and any number of variables might be of interest in a particular study.

Conclusions

Benefits

Clearly, a variety of variables can be extracted from the coordinates acquired by processing the data one time through. Thus, many characteristics of gesture can be determined simultaneously without having to develop separate coding schemes for each variable of interest (i.e., using McNeill-like gesture areas for size and shape, making judgments about speed), although sometimes multiple coding schemes will be necessary. This saves researchers time by not having to make several passes through the data for each variable and obviates the need to make decisions about how to capture each variable separately.

Along similar lines, because this coding scheme describes the form of body movement quantitatively, it is objective and does not require extensive training to implement in a reliable fashion. Typically, form-based properties of movement are described by having coders make a categorical judgment about the size or speed of a movement on a scale (e.g., Holler & Stevens, 2007; Kuhlen et al., 2012) or by identifying movement "areas" post-hoc and evaluating if a movement falls in this region (e.g., Bayliss, 2011; Galati & Brennan, 2013). This requires considerable training to ensure inter-coder reliability. Instead, our coding scheme describes the movement by exploiting the thousands of pixels available on a computer monitor. This reduces the likelihood of coder bias.

The data generated by this method are very reliable. To calculate reliability scores for our trajectory coding, we have been randomly selecting 20 % of all movements for a second coder to process and have then calculated the correlation in the pixel selection between both coders or the distance between points selected by each coder. We have found that variation that occurs is typically very slight (less than 10 pixels). In addition, errant mouse clicks or coder errors are typically easy

to identify because they involve great disparity between coders, and between a single frame and nearby frames of video. Based on our experience, we do not think that reliability would need to be assessed in each application of this technique, but, depending on the size of the studied effect, some researchers may want to further examine variability in their implementation of the technique.

As suggested above, this coding scheme can be readily used in tandem with other coding schemes to maximize the amount of information obtained from video data. This system by no means seeks to replace other well-established coding schemes that identify the bounds of a gesture (Kendon, 2004; McNeill, 1992), describe gesture type (Butterworth & Beattie, 1978; McNeill, 1985), detail semantic features (Beattie & Shovelton, 1999; Gerwing & Allison, 2009), or a number of other aspects of gesture coding. Instead, we hope that this approach will supplement others used by allowing researchers to capture properties of gesture that have been difficult to discriminate, and therefore ignored, in previous research.

Limitations

Despite its benefits, there are also some limitations to our approach. The first, perhaps obvious, limitation is that the motion signal only captures two dimensions: x and y coordinates. Any depth information present in the video data is distorted or entirely ignored. Returning to our example of the gesture depicting the child going down the tornado slide: the spiral motion present could not be captured and analyzed with a single stream of video. This could be corrected for by using an additional camera positioned directly to the side of the participant to capture any depth information. This could then be integrated with the remaining two dimensions. However, this has logistical complications. Although this can be readily done in an experimental setting, one of the major benefits of our measurement system is that it allows for fine-grained analysis of data collected in naturalistic settings. Thus, our coding system may not be of interest to those who have a theoretical interest in depth information, particularly in data that may be collected outside of the lab. We have yet to add an additional camera, as we have found that having just two dimensions has still provided considerable information about how each hand moves.

Another potential limitation, as mentioned earlier, is that this coding scheme necessitates fairly stringent requirements for camera position and angle. This makes interactions in which multiple people are involved and constantly moving more difficult. We have found that in the case of multiple people, positioning cameras slightly above and behind each person pointed toward the other person still allows for clean coding of the video data, capturing gestures as they are seen by an observer, albeit in only two dimensions.

Along the same lines, camera and body position must be relatively stable throughout data collection, or data must be transformed into a common space. This coding scheme may therefore be difficult to employ in naturalistic multi-party or classroom situations, where gesturers may be changing their body positions as they address different addressees. We have yet to use this coding scheme for such data. Though limitations exist, we consider the information that can be gleaned from this coding scheme to outweigh these limitations.

Summary

It is widely accepted that body movement varies across a multitude of dimensions. Although categorical coding schemes have been able to capture some degree of this variation, the richness present in body movements has likely been overlooked due to the limitations in measurement techniques. Our approach provides an accessible and straightforward method that can be incorporated into movement research to help capture this continuous variation. We do not suggest that categorical coding schemes should be avoided, but rather that they be used in concert with continuous measures: having a fine-grained technique for capturing properties of body movement will add greater power to categorical coding and analyses. Even here, in a simple, illustrative experiment, we uncovered new evidence about iconicity in gesture. Using a method that can capture continuous variables will potentially facilitate an understanding of many new aspects of movement. We hope that by sharing these methods, we can help standardize and streamline measurement of movement for researchers from a variety of research traditions.

References

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture–speech integration in narrative: Are children less redundant than adults? *Gesture, 9*(3), 290–311. doi:10.1075/gest.9.3.02ali

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. Retrieved from http://arxiv.org/abs/1406.5823

Bayliss, A. J. (2011). *Impression formation of Chinese EFL speakers: The significance of hands (Unpublished Doctoral Dissertation)*. The University of Melbourne.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*(4), 438–462. doi:10.1177/0261927X99018004005

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. Version 5.4.14. Retrieved 24 July 2015 from http://www.praat.org/

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. *Recent Advances in the Psychology of Language*, 347–360.

Cardona, T. R. (2008). Metaphors in sign languages and in co-verbal gesturing. *Gesture, 8*(1), 62–81.

Chiou, R. Y. C., Wu, D. H., Tzeng, O. J. L., Hung, D. L., & Chang, E. C. (2012). Relative size of numerical magnitude induces a size-contrast effect on the grip scaling of reach-to-grasp movements. *Cortex, 48*(8), 1043–1051. doi:10.1016/j.cortex.2011.08.001

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gesture affect listeners' actions. *Cognition, 113*(1), 98–104. doi:10.1016/j.cognition.2009.06.006

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The Body Action and Posture Coding System (BAP): Development and reliability. *Journal of Nonverbal Behavior, 36*(2), 97–121. doi:10.1007/s10919-012-0130-0

Efron, D. (1972). Gesture, race and culture. The Hague: Mouton.

Frey, S., Hirsbrunner, H. P., Florin, A., Daw, W., & Crawford, R. (1983). A unified approach to the investigation of nonverbal and verbal behavior in communication research. *Current Issues in European Social Psychology, 1,* 143–199.

Galati, A., & Brennan, S. E. (2013). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language and Cognitive Processes*, (January 2014), 1–17. doi:10.1080/01690965.2013.796397

Gerwing, J., & Allison, M. (2009). The relationship between verbal and gestural contributions in conversation: A comparison of three methods. *Gesture, 9,* 312–336. doi:10.1075/gest.9.3.03ger

Goldin-Meadow, S., Mylander, C., & Franklin, A. (2007). How children make language out of gesture: Morphological structure in gesture systems developed by American and Chinese deaf children. *Cognitive Psychology, 55*(2), 87–135. doi:10.1016/j.cogpsych.2006.08.001

Hilliard, C., & Cook, S. W. (2015). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. doi:10.1037/xlm0000154

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology, 26*(1), 4–27. doi:10.1177/0261927X06296428

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495–514. doi:10.3758/PBR.15.3.495

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260–7. doi:10.1177/0956797609357327

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S., Van Gijn, I., & Van Der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In *Gesture and sign language in human-computer interaction* (pp. 23–35). Springer Berlin Heidelberg. Retrieved from doi:10.1007/BFb0052986

Kuhlen, A. K., Galati, A., & Brennan, S. E. (2012). Gesturing integrates top-down and bottom-up information: Joint effects of speakers' expectations and addressees' feedback. *Language and Cognition, 4*(1), 17–41. doi:10.1515/langcog-2012-0002

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*. Retrieved from http://psycnet.apa.org/journals/rev/92/3/350/

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Ochs, E. (1979). Transcription as theory. *Developmental pragmatics, 10*(1), 43–72.

Priesters, M. A. (2013). *Functional patterns in gesture space (Unpublished Doctoral Dissertation)*. Institue für Sprach- und Kommunikationswissenschaft der RWTH Aachen.

Priesters, M., & Mittelberg, I. (2013). Individual differences in speakers' gesture spaces: Multi-angle views from a motion-capture study. *Proceedings of the Tilburg Gesture Research*. Retrieved from http://tiger.uvt.nl/pdf/papers/priesters.pdf

Rodger, M. W. M., Craig, C. M., & O'Modhrain, S. (2012). Expertise is perceived from both sound and body movement in musical performance. *Human Movement Science, 31*(5), 1137–1150. doi:10.1016/j.humov.2012.02.012

Ruiter, J. De. (2000). The production of gesture and speech. In *Language and gesture*. Retrieved from http://pub.uni-bielefeld.de/publication/2529767

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*(4), 585–615. doi:10.1037/0096-3445.112.4.585

Senghas, A., Kita, S., & Ozyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science, 305*(5691), 1779–82. doi:10.1126/science.1100199

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language, 55*(2), 167–177. doi:10.1016/j.jml.2006.03.002

Stokoe, W. C., Casterline, D. C., & Croneberg, C. G. (1976). *A dictionary of American sign language on linguistic principles*. Silver Spring: Linstok Press.

Stungis, J. (1981). Identification and discrimination of handshape in American Sign Language. *Perception & Psychophysics, 29*(3), 261–276. doi:10.3758/BF03207293

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.

Zhao, L., & Badler, N. (2001). Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures. Retrieved from http://repository.upenn.edu/cis_reports/116/

Zhao, L. P., & Kolonel, L. N. (1992). Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies. *American Journal of Epidemiology, 136*(4), 464–474.