Video Article

Frame-by-Frame Video Analysis of Idiosyncratic Reach-to-Grasp Movements in Humans

Jenni M. Karl[1], Jessica R. Kuntz[2], Layne A. Lenhart[2], Ian Q. Whishaw[2]

[1] Department of Psychology, Thompson Rivers University

[2] Department of Neuroscience, University of Lethbridge

Correspondence to: Jenni M. Karl at jkarl@tru.ca

Abstract

Prehension, the act of reaching to grasp an object, is central to the human experience. We use it to feed ourselves, groom ourselves, and manipulate objects and tools in our environment. Such behaviors are impaired by many sensorimotor disorders, yet our current understanding of their neural control is far from complete. Current technologies for investigating human reach-to-grasp movements often utilize motion tracking systems that can be expensive, require the attachment of markers or sensors to the hands, impede natural movement and sensory feedback, and provide kinematic output that can be difficult to interpret. While generally effective for studying the stereotypical reach-to-grasp movements of healthy sighted adults, many of these technologies face additional limitations when attempting to study the unpredictable and idiosyncratic reach-to-grasp movements of young infants, unsighted adults, and patients with neurological disorders. Thus, we present a novel, inexpensive, and highly reliable yet flexible protocol for quantifying the temporal and kinematic structure of idiosyncratic reach-to-grasp movements in humans. High speed video cameras capture multiple views of the reach-to-grasp movement. Frame-by-frame video analysis is then used to document the timing and magnitude of pre-defined behavioral events such as movement start, collection, maximum height, peak aperture, first contact, and final grasp. The temporal structure of the movement is reconstructed by documenting the relative frame number of each event while the kinematic structure of the hand is quantified using the ruler or measure function in photo editing software to calibrate 2 dimensional linear distances between two body parts or between a body part and the target. Frame-by-frame video analysis can provide a quantitative and comprehensive description of idiosyncratic reach-to-grasp movements and will enable researchers to expand their area of investigation to include a greater range of naturalistic prehensile behaviors, guided by a wider variety of sensory modalities, in both healthy and clinical populations.

Video Link

The video component of this article can be found at https://www.jove.com/video/56733/

Introduction

Prehension, the act of reaching to grasp an object, is used for many daily functions including acquiring food items for eating, grooming, manipulating objects, wielding tools, and communicating through gesture and written word. The most prominent theory concerning the neurobehavioral control of prehension, the Dual Visuomotor Channel theory[1,2,3,4], proposes that prehension consists of two movements - a reach that transports the hand to the location of the target and a grasp that opens, shapes, and closes the hand to the size and shape of the target. The two movements are mediated by dissociable but interacting neural pathways from visual to motor cortex via the parietal lobe[1,2,3,4]. Behavioral support for the Dual Visuomotor Channel theory has been ambiguous, largely due to the fact that the reach-to-grasp movement appears as a single seamless act and unfolds with little conscious effort. Nonetheless, prehension is almost always studied in the context of visually-guided prehension in which a healthy participant reaches to grasp a visible target object. Under these circumstances the action does appear as a single movement that unfolds in a predictable and stereotypical fashion. Prior to reach onset the eyes fixate on the target. As the arm extends the digits open, preshape to the size of the object, and subsequently start to close. The eyes disengage from the target just prior to target contact and final grasp of the target follows almost immediately afterwards[5]. When vision is removed, however, the structure of the movement is fundamentally different. The movement dissociates into its constituent components such that an open-handed reach is first used to locate the target by touching it and then haptic cues associated with target contact guide shaping and closure of the hand to grasp[6].

Quantification of the reach-to-grasp movement is most often achieved using a 3 dimensional (3D) motion tracking system. These can include infrared tracking systems, electromagnetic tracking systems, or video based tracking systems. While such systems are effective for acquiring kinematic measures of prehension in healthy adult participants performing stereotypical reach-to-grasp movements towards visible target objects, they do have a number of drawbacks. In addition to being very expensive, these systems require the attachment of sensors or markers onto the arm, hand, and digits of the participant. These are usually attached using medical tape, which can impede tactile feedback from the hand, alter natural motor behavior, and distract participants[7]. As these systems generally produce numerical output related to different kinematic variables such as acceleration, deceleration, and velocity they are also not ideal for investigating how the hand contacts the target. When

using these systems, additional sensors or equipment are required to determine what part of the hand makes contact with the target, where on the target contact occurs, and how the configuration of the hand might change in order to manipulate the target. In addition, infrared tracking systems, which are the most commonly employed, require the use of a specialized camera to track the location of the markers on the hand in 3D space[6]. This requires a direct line of sight between the camera and the sensors on the hand. As such, any idiosyncrasies in the movement are likely to obscure this line of sight and result in the loss of critical kinematic data. There are, however, a large number of instances in which idiosyncrasies in the reach-to-grasp movement are actually the norm. These include during early development when infants are just learning to reach and grasp for objects; when the target object is not visible and tactile cues must be used to guide the reach and the grasp; when the target object is an odd shape or texture; and when the participant presents with any one of a variety of sensorimotor disorders such as a stroke, Huntington's disease, Parkinson's disease, Cerebral Palsy, *etc.* In all of these cases, the reach-to-grasp movement is neither predictable nor stereotypical, nor is it necessarily guided by vision. Consequently, the capability of 3D motion tracking systems to reliably quantify the temporal and kinematic structure of these movements can be severely limited due to disruptions in sensory feedback from the hand, changes in natural motor behavior, loss of data, and/or difficulties interpreting the idiosyncratic kinematic output from these devices.

The present paper describes a novel technique for quantifying idiosyncratic reach-to-grasp movements in various human populations that is affordable, does not impede sensory feedback from the hand or natural motor behavior, and is reliable but can be flexibly modified to suit a variety of experimental paradigms. The technique involves using multiple high-speed video cameras to record the reach-to-grasp movement from multiple angles. The video is then analyzed offline by progressing through the video frames one at a time and using visual inspection to document key behavioral events that, together, provide a quantified description of the temporal and kinematic organization of the reach-to-grasp movement. The present paper describes a comparative analysis of visually- versus nonvisually-guided reach-to-grasp movements in healthy human adults[6,8,9,10] in order to demonstrate the efficacy of the technique; however, modified versions of the technique have also been used to quantify the reach-to-grasp actions of human infants[11] and non-human primates[12]. The comprehensive results of the frame-by-frame video analysis from these studies are among the first to provide behavioral evidence in support of the Dual Visuomotor Channel theory of prehension.

Protocol

All procedures involving human participants have been approved by the University of Lethbridge Human Subjects Research Committee and the Thompson Rivers University Research Ethics for Human Subjects Board.

1. Participants

1. Attain informed consent of adults that have normal or corrected-to-normal vision and are of good health with no history of neurological or sensorimotor disorders (unless the aim is to investigate a particular clinical population).

2. Experimental Setup

1. Select blueberries, donut balls, and orange slices to serve as reaching targets. Measure a subset of ten of each of the targets across their longest axis to determine the mean length of each target.
   NOTE: Utilize targets that are uniform in shape and size. The mean size of the blueberry targets was 12.41 ± 0.33 mm, the mean size of the donut ball targets was 28.82 ± 1.67 mm, and the mean size of the orange slice targets was 60.53 ± 0.83 mm.
2. Determine the trial number and order for the experiment. Inform participants that they will be completing a total of 60 reaching trials separated into 4 blocks (2 blocks in the Vision condition and 2 blocks in the No Vision) with each block consisting of 15 reaching trials (5 repetitions for each of the 3 target objects). Inform the participant that for each block the target objects will be presented in a random order as determined by a random number generator. Ensure that the order of block presentation is counterbalanced across participants.
3. Seat the participant in a stationary armless chair in a quiet, well-lit room free from distractions. Tell the participant to sit up straight in the chair with both feet resting flat and square on the floor and their hands resting open and palm-down on the tops of their upper thighs.
4. Adjust the height of a self-standing, height-adjustable pedestal to the seated participant's trunk length so that the top of the pedestal stands midway between the top of the participant's hip and the participant's sternum. Place the pedestal directly in front of the participant's midline.
5. Tell the participant to extend their dominant hand directly towards the top of the pedestal. Adjust the location of the pedestal so that it is positioned at the participant's midline, but at a distance equivalent to the participant's fully extended arm and hand such that the participant's outstretched middle finger contacts the distal edge of the pedestal. After positioning the pedestal, ask the participant to return their outstretched hand to their lap.
6. Position 1 high speed video camera sagittal to the participant, on the same side as the participant's non-dominant hand at a 1 m distance from the pedestal to record a reach-side view of the participant's dominant hand. Adjust the position and zoom of the camera until the top of the participant's head, the start position of the hand on the upper thigh, and the reaching target on the pedestal are all clearly visible from this camera angle.
7. Position a second video camera 1 m in front of the pedestal to capture a front-on view of the participant. Adjust the position and zoom of the camera until the top of the participant's head, the start position of the hand on the thigh, and the reaching target on the pedestal are all clearly visible from this camera angle.
   NOTE: Additional video cameras may be positioned above, below, or in front of the participant and pedestal as desired.
8. Set each camera to record video at the highest possible resolution at a rate of 60, 120, or 300 frames per second with a shutter speed of 1/250th (or up to 1/1000th if the movement will be performed very quickly) of a second. Set both cameras to store each video file as either an AVI, MP4, or MOV file. Use a strong lamp containing cool LED lights (which generate negligible heat) to illuminate the participant and testing area. Set each camera to focus on the center of the pedestal.
   NOTE: At these high frame rates and shutter speeds a strong lamp is needed to illuminate the participant and testing area. This will ensure that the individual video frames are sufficiently illuminated and free of motion artifacts.
9. Instruct the participant to begin each reaching trial with their hands open, relaxed, and resting palm down on the dorsum of their upper thighs.

10. Tell the participant that at the beginning of each trial, the experimenter will place a target object – either a blueberry, a donut ball, or an orange slice – on the pedestal and that the participant is to wait until the experimenter provides a verbal '1, 2, 3, Go' command to reach out with their dominant hand, grasp the target object, and then place the target object in their mouth as if they are going to eat it.

11. Tell the participant that they should perform the task as naturally as possible but that they do not actually have to eat the target object. Instruct the participant that after placing the target object in the mouth they should then use their non-dominant hand to retrieve the target from their mouth and place it in a disposal container located on the floor adjacent to the participant's non-dominant hand. Instruct the participant to then return both hands to the start position on their upper thighs in preparation for the next trial.

12. Select a blindfold that is not cumbersome but does occlude both foveal and peripheral vision of the target. Provide this blindfold to all participants at the beginning of all No Vision trial blocks and ensure they wear it before the target object is placed on the pedestal.
NOTE: When completing No Vision trial blocks, participants are blindfolded **before** the first No Vision trial begins. Thus, they are blindfolded before a target object is placed on the pedestal which ensures that the participant does not see which of the possible target objects is placed on the pedestal for any given No Vision trial.

13. Press the 'record' button on both video cameras **before** initiating the experiment and ensure that the position and location of each camera does not change for the duration of the experimental task for a given participant.

3. Data Collection

1. Begin the experiment by quickly tapping the top central surface of the pedestal with your index finger.
NOTE: The moment of contact between your index finger and the pedestal will serve as a time cue that will be visible in all video records.

2. Place a calibration object of known size, such as a 1 cm^3 plastic cube, at the center of the top of the pedestal such that each camera has a fronto-parallel view of one side of the cube. Leave the calibration object on the pedestal for approximately 5 s so that each video camera captures an unobstructed view of it, then remove the calibration object before the first reaching trial.

3. Inform the participant that the experiment is about to start, ensure that the participant is wearing the blindfold if they are about to complete a No Vision trial block, and ask the participant to verbally confirm if they are ready to begin.

4. Place the first target object on the pedestal and use a "1, 2, 3, Go" cue to signal to the participant to perform the reaching trial.

5. Repeat step 3.4. until the participant has completed a total of 60 reaching trials. Ensure that the participant only wears the blindfold for the No Vision trial blocks.

6. After 60 reaching trials are complete, stop recording from the video camera. Answer any final questions that the participant may have and allow them to leave.

4. Prepare Videos for Frame-by-Frame Video Analysis

1. Download the video files from the video camera to a secure computer that has a video editing software program installed on it.

2. Open the video files in the video editing software program. In the **Start Window** that opens, click on the **New Project** button. For the **Video Display Format** option select **Frames**. For the **Capture Format option** select **DV**. Click **OK | Yes**.

3. Click **Media Browser** tab and navigate to find the video files for your participant. Click and hold one of the video files to drag and drop it in the adjacent **Timeline**. This will cause the video record to appear in the **Program** window. Use the arrow keys on the keyboard to progress forward and/or backward through the video record.

4. Use the arrow keys on the keyboard to navigate to the video frame that depicts the moment the experimenter taps the top of the pedestal with her index finger. Pause the video record on this frame so that the playhead (on the timeline) is positioned at the exact frame where the experimenter's finger first makes contact with the pedestal.

5. Use the trim function in the video editing software to trim (remove) all frames prior to the current frame. To do this, click the **Mark In | File | Export | Media** options. In the **Export Settings** window that opens, select **H.264** for the format option and **Match Source** for the preset option.

6. Click on the **Output Name** and browse to find the folder where you would like to save the newly trimmed video record. Provide a new file name for the newly trimmed video record that you are creating, then click the **Save** button. This will return you to the **Export Settings**. Click the **Export** option.

7. Repeat steps 4.1-4.6 for all of the video records for each participant to create a newly trimmed video file that corresponds to each of the original video files. Only use the newly trimmed video files for all subsequent frame-by-frame video analyses.
NOTE: In the newly trimmed video files, frame 1 of each video file will depict the same behavioral event (*e.g.*, moment of first contact between the experimenter's finger and the pedestal) and are in essence time-synchronized. This allows for quick and easy switching between different video views of a single behavioral event within a single testing session for a single participant.

8. Close and re-open the video editing software. Repeat steps 4.2. and 4.3. Select and drag all of the newly trimmed video records for a single participant into separate timelines in the video editing software for the frame-by-frame video analysis. This will allow you to navigate through the multiple video views for each participant in a time-synchronized manner. To change which video record (*e.g.*, front or side) is displayed in the **Program** window, simply click and drag the video timeline that contains the preferred video view to the top of the other video timelines.
NOTE: Step 4.8. is conducted using the video editing software and serves to temporarily time-synchronize all of the video views from a single participant.

5. Frame-by-Frame Video Analysis: Temporal Organization

1. For each reaching trial, describe the temporal organization of the reach-to-grasp movement using the arrow keys on the keyboard to progress through the time-synchronized video records frame-by-frame. Record, in a spreadsheet (**Supplemental Table 1**), the first frame number for each key behavioral event described in steps 5.1.1-5.1.6, which are also described in **Table 1** and illustrated in **Figure 1**.
NOTE: While all 6 key behavioral events are generally present in every Vision trial, some may not always be present in the No Vision trials.
    1. Identify movement start, which is defined as the first visible lifting of the palm of the hand away from the dorsum of the upper thigh.
    2. Identify collection, which is defined as the formation of a closed hand posture in which the digits maximally flex and close. Generally, collection occurs following movement start and prior to peak aperture.

3. Identify maximum height, which is defined as the maximum height of the most proximal knuckle of the index finger as the hand reaches towards the target object.
4. Identify peak aperture, which is defined as the maximum opening of the hand (as measured between the central tip of the index finger and the central tip of the thumb) that occurs after collection but prior to first contact. Sometimes the digits will re-open after first contact with the target object, in which case, also record the frame number of this second peak aperture.
5. Identify first contact, which is defined as the first point of contact between the hand and the target object.
6. Identify final grasp, which is defined as the moment at which all manipulation of the target object is complete and the participant has a firm hold on the target object.

6. Frame-by-Frame Video Analysis: Kinematic Calibration Scale

1. Create a calibration scale for each participant that can be used to convert distance measures taken from the video record from pixels to centimeters.
    1. Drag and drop the video record of interest into the timeline of the video editing software program as in steps 4.2. and 4.3. Move the playhead to the frame that depicts the calibration object and click the **Export Frame**. In the **Export Frame** window that opens, enter a name for the still frame image in the **Name** option box, enter **TIFF** into the **Format** option box, and click on the **Path** option box to browse to the folder that you would like to save the still frame image in.
    2. Open this still frame image file in a photo editing software program. Click **Image | Analysis | Set Measurement Scale | Custom** to transform the mouse pointer into a ruler tool. Use the ruler tool to click on one side of the 1 cm^3 calibration cube, drag the ruler tool to the opposite side of the calibration cube, keeping the line as horizontal as possible, and releasing the click at the opposite side of the cube.
    NOTE: Once step 6.1.2. is complete the photo editing software program will automatically compute the length of the line that was drawn in pixels and display this value in the **Pixel Length** option in the opened **Measurement Scale** window.
    3. In the **Measurement Scale** window, enter **10** into the **Logical Length** option box and **millimeters** into the **Logical Units** option box. Click the **Save Preset**. In the **Measurement Scale Preset** window, enter the video view and code/number of the relevant participant (*e.g.*, **SideView-Participant1**) into the **Preset Name** option box and then click **OK**.
    4. Click **OK** in the **Measurement Scale** window.
    NOTE: Repeat steps 6.1.1 to 6.1.4. for each video view for each participant.

7. Frame-by-Frame Video Analysis: Kinematic Structure

1. For each reaching trial, describe the kinematic structure of the reach-to-grasp movement using the ruler tool in the photo editing software program to record the relevant distance measures described in steps 7.4-7.9 and **Table 1**.
2. Use the video editing software to export a still frame image (step 6.1.1.) that depicts each of the following behavioral events: collection, maximum height, peak aperture, first contact, and final grasp (for each trial).
3. Open the still frame image that depicts the key behavioral event of interest in the photo editing software. Click **Image | Analysis | Set Measurement Scale** and select the preset calibration scale that corresponds to the video view and participant depicted in the image that you wish to take a distance measurement from (*e.g.*, **SideView-Participant1**).
    NOTE: Selecting the appropriate preset calibration scale will ensure that all subsequent distances measured with the ruler tool will be accurately converted from pixels into millimeters. The preset calibration scale will remain automatically selected for all subsequent image files that are opened. Thus, there is no need to repeat step 7.3. until you switch to analyzing still frame images from a different video view or a different participant.
4. Open the still frame image that depicts the key behavioral event of collection in the photo editing software. Select the ruler tool and use it to draw a straight line between the central tip of the thumb and the central tip of the index finger.
5. Click **Image | Analysis | Record Measurements**, which will cause the **Measurement Log** to open. Record the **Length** of this line as the **collection distance** in the spreadsheet (**Supplemental Table 1**).
6. Open the still frame image that depicts maximum height in the photo editing software. Use the ruler tool to measure the vertical distance between the top of the pedestal and the top of the participant's index knuckle. Record the **Length** of this line as the **maximum height distance** in the spreadsheet.
7. Open the still frame image that depicts peak aperture in the photo editing software. Use the ruler tool to measure the distance between the central tip of the thumb and the central tip of the index finger. Record the **Length** of this line as the **peak aperture distance** in the spreadsheet.
8. Open the still frame image that depicts first contact in the photo editing software. Use the ruler tool to measure the distance between the central tip of the thumb and the central tip of the index finger. Record the **Length** of this line as the **first contact aperture distance** in the spreadsheet.
9. Open the still frame image that depicts final grasp in the photo editing software. Use the ruler tool to measure the distance between the central tip of the thumb and the central tip of the index finger. Record the **Length** of this line as the **final grasp aperture distance** in the spreadsheet.

8. Frame-by-Frame Video Analysis: Topographical Measures

1. While performing the above frame-by-frame video analysis, also document additional topographical features of the reach-to-grasp movement such as part of hand to make first contact, contact points, grasp points, adjustments, grip type, and grasp strategy (**Table 2**).
    1. Document, in the spreadsheet, which part of the hand is used to make first contact with the target for each trial, for each participant. Use the following notation: 1 = thumb, 2 = index finger, 3 = middle finger, 4 = ring finger, 5 = pinky, 6 = palm, 7 = dorsum of the hand.

2. Determine first contact points by exporting a still frame image of the target, opening it in the photo editing software, and using the program's paintbrush tool to mark the location on the target at which first contact between the hand and the target was made for each trial. Adjust the size, opacity, and color of the paintbrush tool to suit your needs. Repeat this step until you have created a single topographical map that indicates the location of first contact points on the target for each participant.
NOTE: For an example of aggregated first contact points across all of the participants in a single study, see Representative Results below.

3. Determine the grasp points by exporting a still frame image of the target, opening it in the photo editing software, and using the program's paintbrush tool to mark the location on the target at which the hand contacts the target at the time of final grasp for each trial. Adjust the size, opacity, and color of the paintbrush tool to suit your needs. Repeat this step until you have created a single topographical map that indicates the location of grasp points on the target for each participant.
NOTE: For an example of aggregated grasp points across all of the participants in a single study, see Representative Results below.

   1. Visually determine the average grasp contact locations for the thumb and the opposing digit on the target for sighted participants. Denote these two contact locations as the "baseline grasp contact points"
   2. Use the paintbrush tool to mark the "baseline grasp contact points" on the topographical map that illustrates first contact points for each participant. Then use the ruler tool (see steps 6.1. to 6.1.4. and 7.5.) to measure the 2D linear distance between each first contact point and the respective baseline contact point. Repeat this step for every first contact point for each participant in the vision and the no vision conditions. Calculate the average "distance to baseline contact point" for each participant, which will indicate how far, on average, a participant's location of first contact differed from the baseline grasp contact point.
   3. Use the paintbrush tool to mark the "baseline contact points" on the topographical map that illustrates grasp contact points for each participant. Then use the ruler tool (see steps 6.1. and 6.1.4. and 7.5.) to measure the 2D linear distance between each grasp point and the respective baseline contact point. Repeat this step for every grasp point for each participant in the vision and no vision conditions. Calculate the average "distance to baseline contact point" for each participant, which will indicate how far, on average a participant's grasp contact points differed from the baseline grasp contact point.

4. Determine the number of adjustments made on each trial by inspecting the video record, noting any instances where the participant released and re-established contact with the target between the frame of first contact and the frame of final grasp. Record the total number of adjustments per trial for every participant in the spreadsheet.

5. Determine the grip type used to pick up the target for each trial and record it in the spreadsheet: (i) pincer grip: characterized by gripping the target between the pads of the thumb and one other digit of the same hand, (ii) precision grip: characterized by gripping the target between the pads of the thumb and at least two other digits of the same hand, or (iii) power grip: characterized by gripping the target between the palm and the digits of the same hand.

6. Determine the grasp strategy (preshaping, touch-then-grasp, variation 1, variation 2, or variation 3 strategy; see Representative Results below) used for each trial and record it in the spreadsheet.

Representative Results

This section provides examples of the results that can be obtained when using frame-by-frame video analysis to investigate idiosyncratic reach-to-grasp movements under nonvisual sensory guidance. The primary finding is that when participants can use vision to preemptively identify both the extrinsic (location/orientation) and intrinsic (size/shape) properties of a target object they integrate the reach and the grasp into a single seamless prehensile act in which they preshape the hand to the size and shape of the target before touching it (**Figure 2A**). When vision is unavailable, however, they dissociate the two movements so that tactile feedback can be used to first direct the hand in relation to the extrinsic and then in relation to the intrinsic properties of the target, in what has been termed a generalized touch-then-grasp strategy (**Figure 2B**). The results derived from frame-by-frame video analysis are comparable to that of traditional motion tracking systems without the expense, hassle, and other drawbacks of attaching sensors to the participant's hands. The results also provide support for the postulate of the Dual Visuomotor Channel Theory of Prehension that the reach and the grasp are separable movements that appear as one when integrated together under visual guidance.

All key behavioral events are generally present in both the Vision and No Vision Conditions. However, there is a noticeable change in the No Vision condition, such that a significantly greater amount of time is required to transition from peak aperture to first contact and again from first contact to final grasp (**Figure 3**). Review of the kinematic results from the frame-by-frame video analysis provides a number of explanations for this increase in movement duration in the No Vision condition.

The hand takes a more elevated approach to the target and thus, achieves a greater maximum height in the No Vision condition compared to the Vision condition (**Figure 4**). This greater maximum height is a consistent feature of the No Vision reach-to-grasp movement, even after 50 trials of practice. The use of a more elevated reaching trajectory, in which the hand is raised above the target and then lowered down onto it from above, likely contributes to the increased amount of time required to transition from peak aperture to first contact in the No Vision compared to Vision conditions.

In the No Vision condition, the hand maintains a neutral posture, in which the digits remain open and extended during transport towards the target. This differs from the Vision condition in which the digits flex and close into a configuration that matches the size of the target on approach towards it. Consequently, in the No Vision condition the aperture of the hand does not preshape to the size of the target at either peak aperture (**Figure 5**, top) or at first contact (**Figure 5**, middle). This lack of preshaping in the No Vision condition means that additional time is required to modify the hand's configuration after first contact in order to match that of the target. This contributes to the increased amount of time required to transition from first contact to final grasp in the No Vision condition. Despite differences in hand aperture prior to and at first contact with the target, hand aperture at final grasp is identical in the Vision and No Vision conditions (**Figure 5**, bottom).

In the No Vision Condition, the location at which the thumb (red) or index finger (blue) made first contact with the target is randomized across the dorsal surface of the target object in a haphazard way, indicating the absence of a preferred digit-thumb orientation (**Figure 6**, bottom left). This differed from the Vision condition in which the index finger and thumb consistently established first contact with opposing sides of the target, indicating the presence of a preferred digit-thumb orientation prior to first contact (**Figure 6**, top left). The absence of a preferred digit-thumb orientation prior to first contact in the No Vision condition meant that additional time was required after first contact to re-adjust the configuration and position of the digits towards appropriate grasp points that were conducive to actually grasping the target. This is eventually achieved by the time of final grasp (**Figure 6**, bottom right) with a consistency similar to that observed in the Vision condition (**Figure 6**, top right).

In the No Vision condition, participants generally make at least one adjustment after first contact with the target (**Figure 7**), usually to re-direct the digits to more appropriate grasp points on the target. In contrast, in the Vision condition, participants never adjust hand-to-target contact after first contact. Thus, the adjustments made by participants in the No Vision condition likely contribute to the increased amount of time required to transition from first contact to final grasp.

Figure 8 illustrates the part of the hand used to make first contact with the target in the Vision condition (**Figure 8A**, left) and in the No Vision Condition (**Figure 8B**, left). In the Vision condition, participants generally use the index finger and/or thumb to make first contact with the target. In contrast, the part of the hand to make first contact with the target is much more variable in the No Vision condition, with participants often using any of the digits or the palm to make first contact. Notably, in the Vision condition the digits to make first contact with the target are the same ones that make contact during the final grasp. In contrast, the parts of the hand used to make first contact in the No Vision condition are usually different from the parts of the hand used during the final grasp (**Figure 8A** & **Figure 8B**, right).

Figure 9 illustrates the proportion of trials on which participants used a pincer or precision grasp to acquire the target object. Participants in the No Vision condition used a precision grip significantly more than a pincer grip, in contrast to participants in the Vision condition, who preferred a pincer grip.

In the Vision condition, participants consistently use a preshaping strategy in which the hand shapes and orients to the target prior to first contact in order to facilitate immediate grasping of the target. In the No Vision condition, the hand does not shape or orient to the target prior to first contact. Rather, in the No Vision condition the preferred grasp strategy is a touch-then-grasp strategy. This strategy is characterized by initial contact with the target, followed by a release of contact during which the hand re-shapes and re-orients, resulting in altered digit-to-target contact locations that ultimately facilitate successful grasping of the target (**Figure 10A**). Depending on the configuration of the hand at the time of first contact, variations of the touch-then-grasp strategy could be observed. In the first variation (**Figure 10B**), the hand is semi-shaped at first contact and first contact is made with the index finger or thumb, but at an inappropriate contact location, resulting in modifications in both hand shape and contact location prior to establishment of the final grasp posture. In the second variation (**Figure 10C**), the hand does not shape at all prior to first contact, but first contact is made with an appropriate part of the hand at an appropriate location on the target. Thus, a simple flexion of the remaining digits allows for successful capture of the target between the digits and thumb in an effective grasping posture. In the third variation (**Figure 10D**), the hand does not shape at all prior to first contact and first contact is made at an inappropriate location on the target, but with an appropriate part of the hand. Thus, the digit that makes first contact maintains contact while adjacent digits manipulate the target into a position that more readily facilitates grasping of the target between the index/middle finger and the thumb.
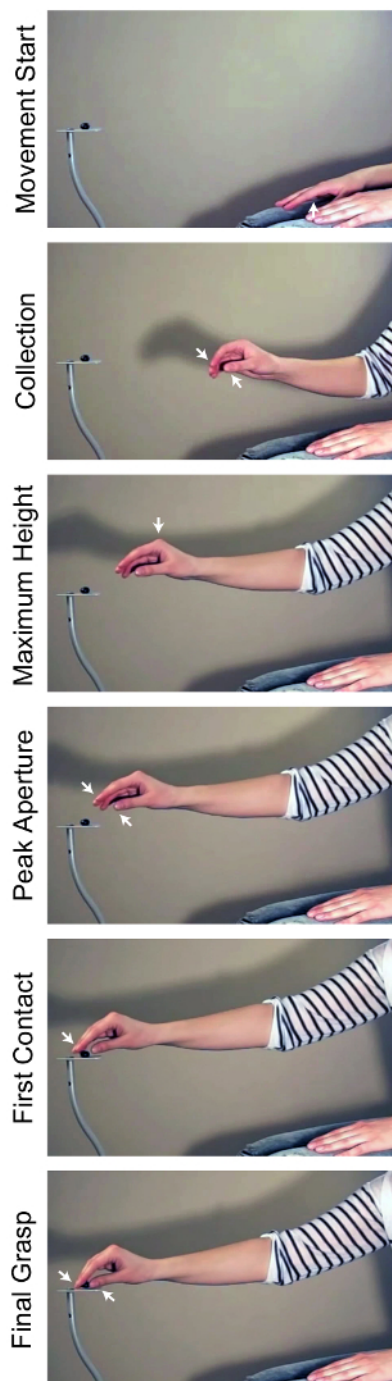
Figure 1: Six behavioral events. Still frames illustrating the 6 key behavioral events that constitute a stereotypical visually-guided reach-to-grasp movement in healthy human adults. White arrows indicate the aspects of the hand/action that are most relevant for identifying each behavioral event. Participants reached with their dominant hand. Please click here to view a larger version of this figure.
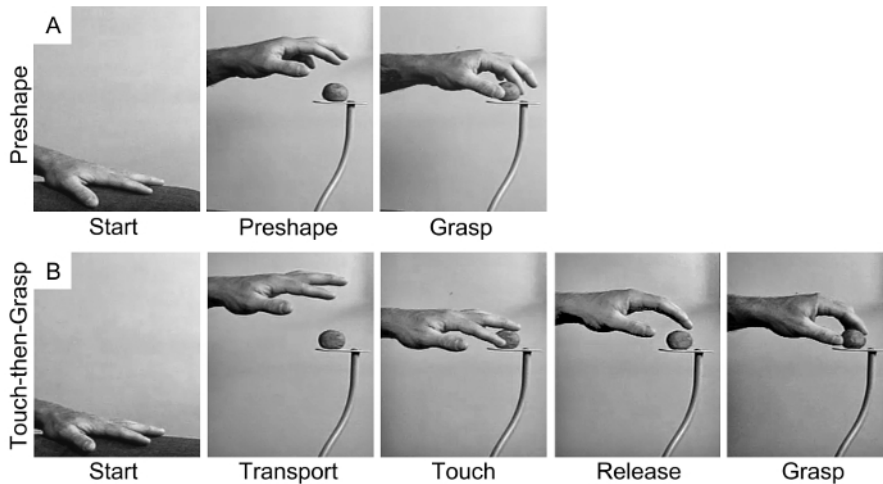
Figure 2: Grasping strategies used by adults in the Vision and No Vision conditions. Still frames illustrating the preshaping strategy (**A**) that was favored by participants in the Vision condition and the general touch-then-grasp strategy (**B**) that was favored by participants in the No Vision condition. Participants reached with their dominant hand. This figure has been modified from Karl *et al.*[6] and Whishaw *et al.*[11] Please click here to view a larger version of this figure.
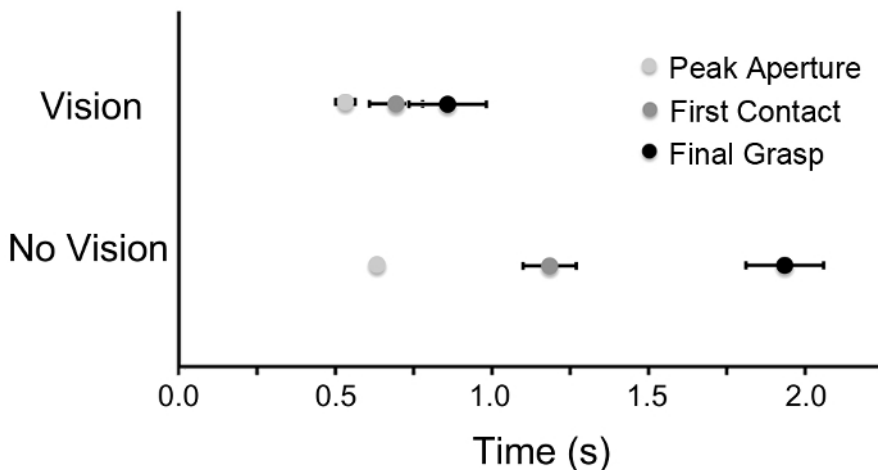

Figure 3: Temporal organization of the reach-to-grasp movement. Time (mean ± standard error (SE)) to peak aperture (light grey), first contact (medium grey), and final grasp (black) of the reach-to-grasp movement of participants (n = 12) in the Vision (top) and No Vision conditions (bottom). This figure has been modified from Karl *et al.*[6] Please click here to view a larger version of this figure.
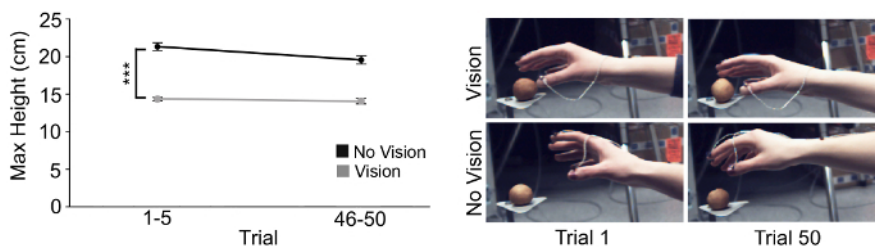

Figure 4: Maximum height. Maximum height (mean ± SE) of the reach-to-grasp trajectory for the first five versus the last five trials of each participant (n = 20) in the Vision and No Vision conditions (A). These results were confirmed by a repeated measures analysis of variance (ANOVA) that found a main effect of Condition $F(1,17) = 35.673$, $p < 0.001$ but no main effect of Trial $F(9,153) = 1.173$, $p > 0.05$ (*** = $p < 0.001$). Representative still frames of the arm and hand at the moment of maximum height on the first and last experimental trials in the Vision and No Vision condition (B). Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl *et al.*[8] Please click here to view a larger version of this figure.
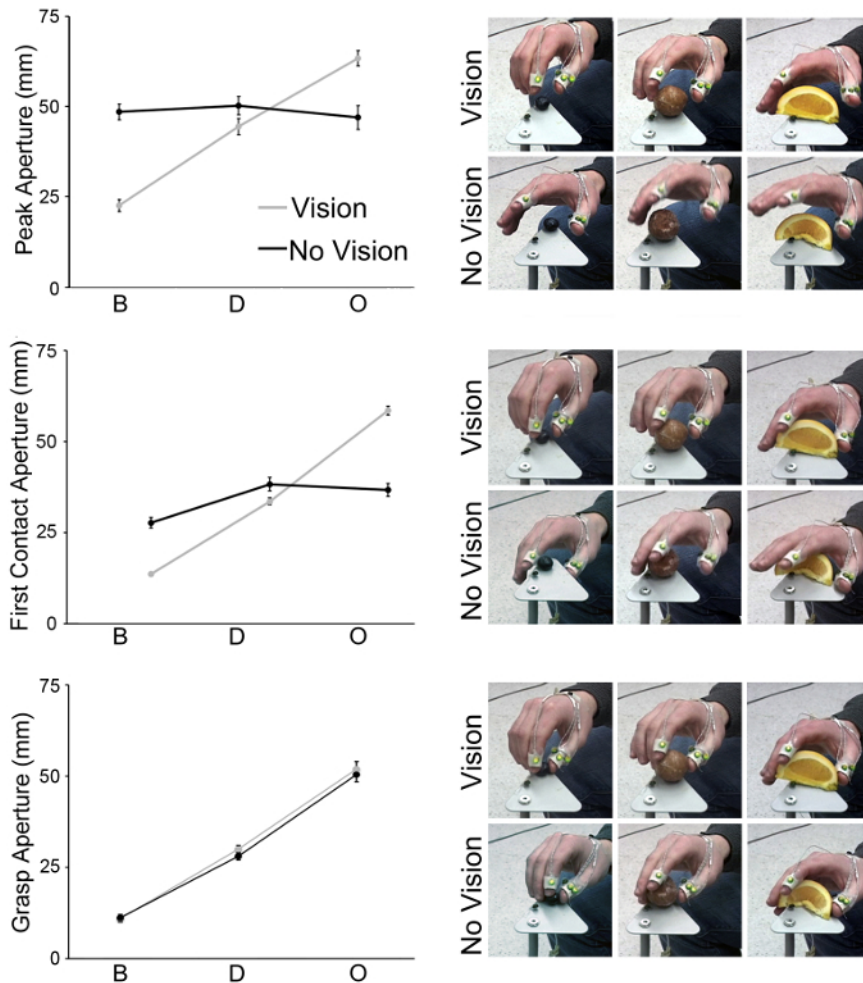
Figure 5: Aperture. Peak aperture (mean ± SE; top), aperture at first contact (mean ± SE; middle), and aperture at final grasp (mean ± SE; bottom) of participants (n = 12) reaching in the Vision (gray) and No Vision (black) conditions. These results were confirmed by repeated measures ANOVAs that found a significant Condition X Target interaction for peak aperture $F(2,20) = 101.088$, $p < 0.001$ and aperture at first contact $F(2,20) = 114.779$, $p < 0.001$, but not for aperture at final grasp $F(2,20) = 0.457$, $p > 0.05$ (*** = $p < 0.001$). Note, the aperture measures shown in the graphs were derived using both a traditional 3D motion tracking system and frame-by-frame video analysis. Participants reached with their dominant hand. B = blueberry, D = donut ball, O = orange slice. This figure has been modified from and presents data originally published in Karl *et al.*[6] Please click here to view a larger version of this figure.
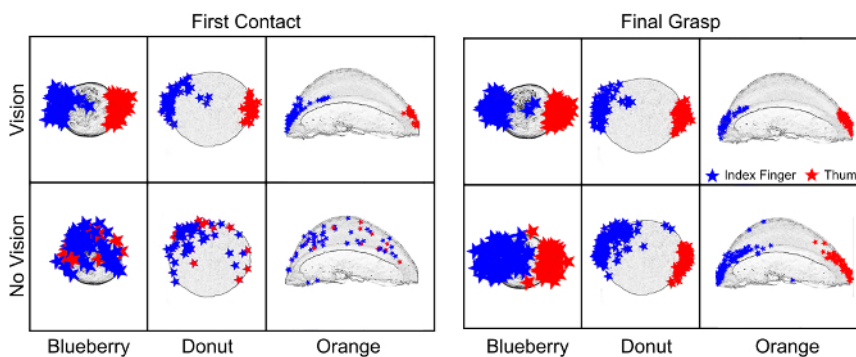

Figure 6: First contact points and grasp contact points. Location of contact points at the moment of first contact with the target (left) and final grasp of the target (right). Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl *et al.*[6] Please click here to view a larger version of this figure.
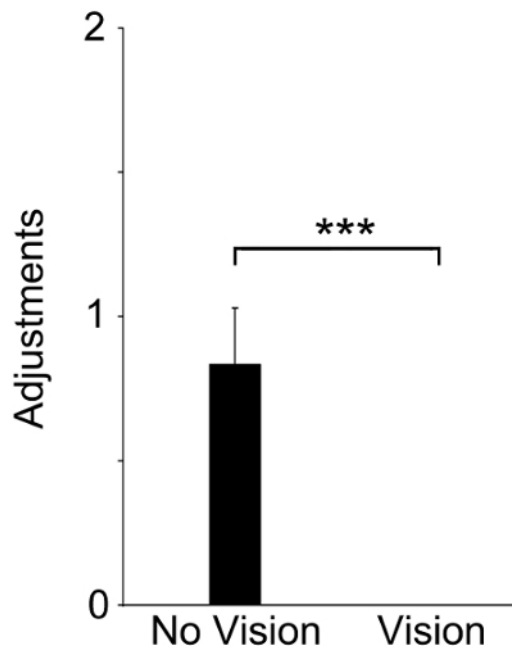
Figure 7: Adjustments. Number of adjustments (mean ± SE) between first contact and final grasp for all participants (n = 18) in the No Vision and Vision conditions. These results were confirmed by a repeated measures ANOVA that gave a significant effect of Condition $F(1,17) = 55.987$, $p < 0.001$ (*** = $p < 0.001$). Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl et al.[10] Please click here to view a larger version of this figure.
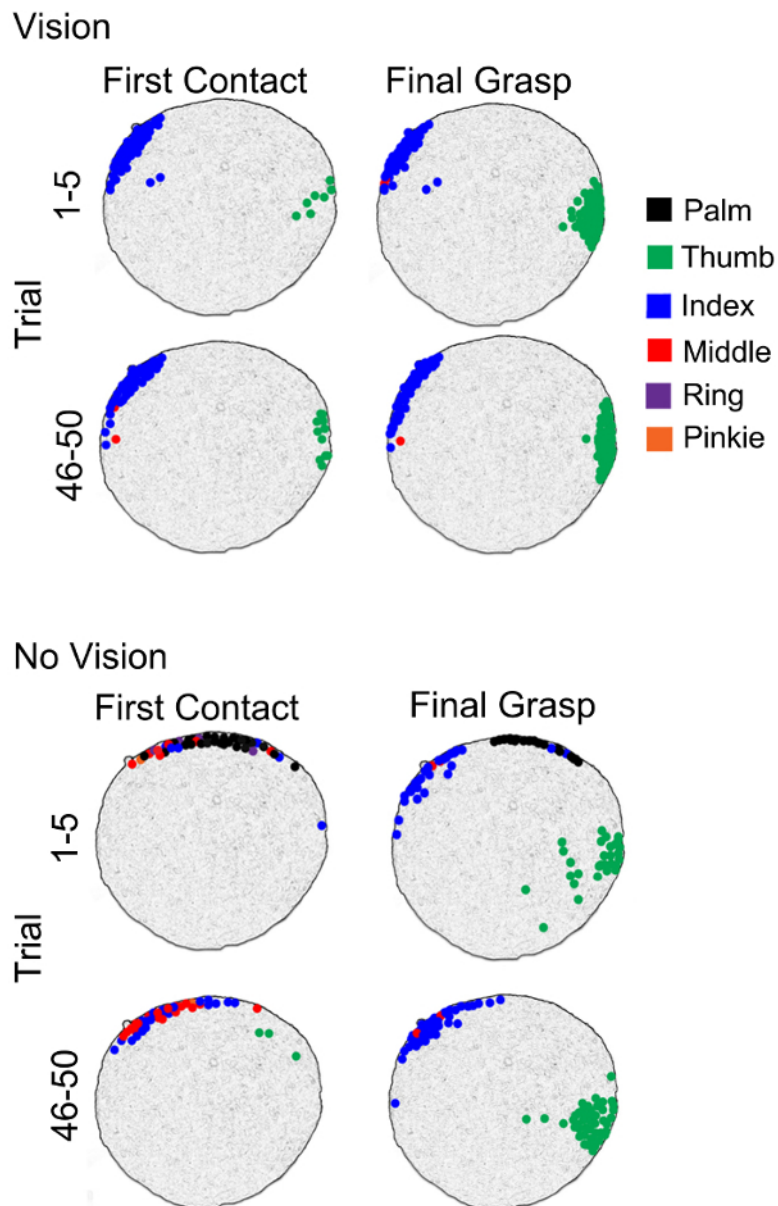
Figure 8: Part of the hand to make contact with the target. Part of the hand to make first contact (left) and final grasp contacts (right) with the target object on the first five and last five experimental trials in the Vision (top) and No Vision (bottom) conditions. Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl *et al.*[8] Please click here to view a larger version of this figure.
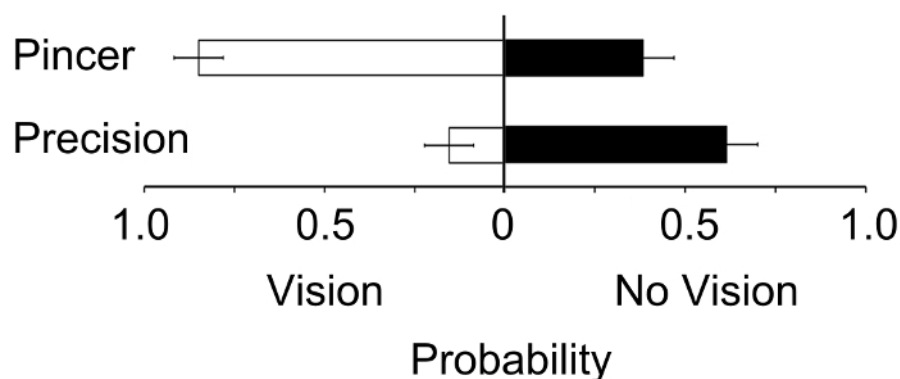
Figure 9: Grip type: Proportion of trials (mean ± SE) for which the participants (n = 12) utilized either a pincer or precision grip to acquire the target in the Vision (A) and No Vision (B) conditions. These results were confirmed by a repeated measures ANOVA that found a significant effect of Condition X Grip $F(1,11) = 32.301$, $p < 0.001$ (*** = $p < 0.001$). Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl et al.[6] Please click here to view a larger version of this figure.
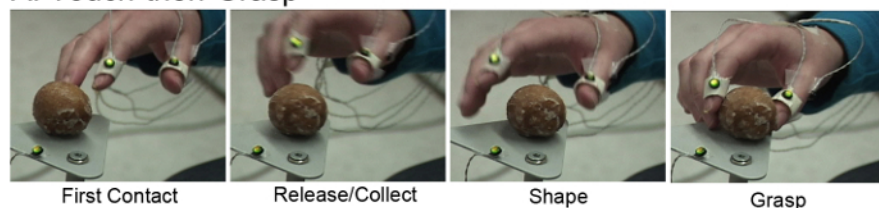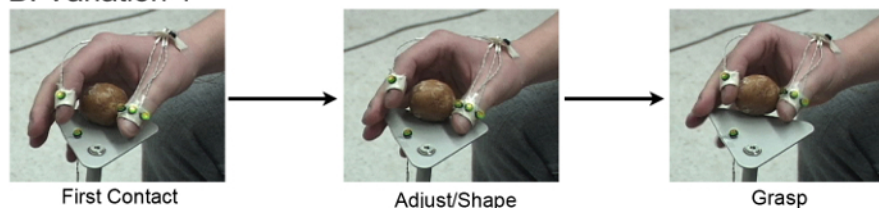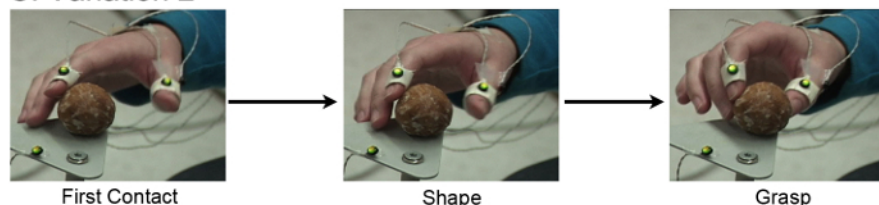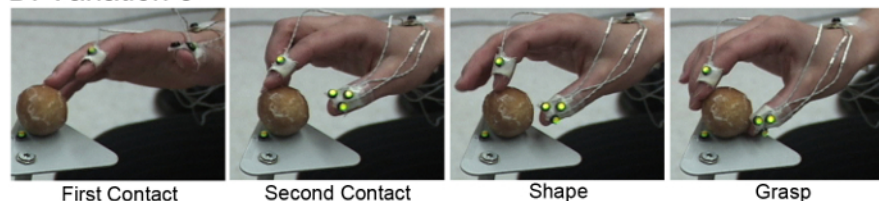
Figure 10: Grasping strategies. Representative still frames illustrate the general touch-then-grasp strategy (**A**), as well as 3 variations of it (**B-D**) by participants in the No Vision condition. Participants reached with their dominant hand. This figure has been modified from and presents data originally published in Karl et al.[6] Please click here to view a larger version of this figure.

| Key Behavioral Event	Description	Record
1. Movement Start	Defined as the first visible lifting of the palm of the hand away from the dorsum of the upper thigh	> Frame number
2. Collection	Defined as the formation of a closed hand posture in which the digits maximally flex and close. Collection may be very obvious or very subtle	> Frame number > Distance between the central tip of the index finger and the central tip of the thumb
3. Maximum Height	Defined as the maximum height of the most proximal knuckle of the index finger	> Frame number > Vertical distance between the top of the pedestal and the top of the index knuckle
4. Peak Aperture	Defined as the maximum opening of the hand, as measured between the two digits used to secure the final grasp of the object, usually the index finger and thumb. In some cases the digits will re-open after target contact and it will be necessary to record a second peak aperture after target contact	> Frame number > Distance between the central tip of the index finger and the central tip of the thumb
5. First Contact	Defined as the moment of first contact between the hand and the target	> Frame number > Distance between the central tip of the index finger and the central tip of the thumb > Part of the hand to make first contact with the target (**Figure 8**) > First contact points (**Figure 6**)
6. Final Grasp	Defined as the moment at which all manipulation of the target is complete and the participant establishes a firm hold on the target	> Frame number > Distance between the central tip of the index finger and the central tip of the thumb > Grasp contact points (**Figure 6**) > Grip type > Part of the hand to make contact with the target at final grasp (**Figure 8**)

Table 1: Description of key behavioral events. Lists the 6 key behavioral events that can be acquired using frame-by-frame video analysis (first column). Each behavioral event is accompanied by a description (second column) as well as a list of the temporal and kinematic information that should be recorded for each (third column).

Topographical Measure	Description	Record
Part of Hand to Make First Contact	Describes what part of the hand was used to make first contact with the target (1= thumb, 2 = index finger, 3 = middle finger, 4 = ring finger, 5 = pinky finger, 6 = palm, 7 = dorsum of hand)	> Which part of the hand was used to make first contact with the target
Contact Points	Illustrates where on the target first contact with the hand occurred	> See step 8.1.2.
Grasp Points	Illustrates where on the target the hand made contact while establishing final grasp of the target	> See step 8.1.3.
Adjustments	A reach-to-grasp movement is considered to contain an adjustment if, between first contact and final grasp, the participant releases and re-establishes contact with the target	> Number of adjustments per trial
Grip Type	Describes the grip configuration used to acquire the target object	> See step 8.1.5.
Grasp Strategy	Refers to the use of different digit-to-target manipulations after first contact in order to facilitate successful grasping of the target	> Type of grasp strategy used (**Figure 10**)

Table 2: Description of topographical measures. Lists the topographical measures that can be acquired using frame-by-frame video analysis (first column). Each measure is accompanied by a description (second column) as well as a list of the types of information that should be recorded for each (third column).

Supplemental Table 1: Spreadsheet for data collection. A template for organizing the temporal, kinematic, and topographical measures (not including contact points and grasp points) collected from frame-by-frame video analysis in a single spreadsheet. Please click here to download this file.

Discussion

The present paper describes how to use frame-by-frame video analysis to quantify the temporal organization, kinematic structure, and a subset of topographical features of human reach-to-grasp movements. The technique can be used to study typical visually-guided reach-to-grasp movements, but also idiosyncratic reach-to-grasp movements. Such movements are difficult to study using traditional 3D motion tracking systems, but are common in developing infants, participants with altered sensory processing, and patients with sensorimotor disorders such as blindness, Parkinson's disease, Stroke, or Cerebral Palsy. Thus, the use of frame-by-frame video analysis will allow researchers to expand their area of investigation to include a greater range of prehensile behaviors, guided by a wider variety of sensory modalities, by both healthy and clinical populations. Specific advantages of frame-by-frame video analysis include its relative affordability, ease to implement, lack of sensors or markers that hinder sensory and motor abilities of the hands, compatibility with other motion tracking systems, and the ability to describe subtle changes in the reach-to-grasp movement that are often hard to interpret from the kinematic output provided by most traditional 3D motion tracking systems. Together, these features of frame-by-frame video analysis have made it possible to advance our theoretical understanding of the neurobehavioral control of prehension.

While there are many instances in which frame-by-frame video analysis may be the only reliable option for analyzing idiosyncratic reach-to-grasp movements, it is important to note that the technique does face some limitations. First, the distance measures (*e.g.*, peak aperture) acquired using frame-by-frame video analysis are 2D and less precise compared to traditional 3D motion tracking systems. Nonetheless, if necessary, additional cameras could be focused on the region of interest. This would allow the experimenter to select the camera view that provides the clearest fronto-parallel view of the behavioral event of interest, and thus increase the precision of the distance measure for that particular event. Furthermore, if very high precision is required for the distance measures then frame-by-frame video analysis can easily be combined with traditional 3D motion tracking techniques (see **Figure 4**, **5**, and **10**) as it does not impede data collection from the traditional system. Secondly, the ultimate success of the technique is critically dependent on the integrity of the video record. Choosing filming views that adequately capture the behavior, using a shutter speed of 1/1000th of a second with a strong light source, and ensuring that the focus of the camera remains stabilized on the action of interest will all help to ensure that individual frames in the video record are crisp, free of motion artifacts, and easy to analyze. Finally, when first learning to implement the technique, researchers may wish to utilize multiple blind raters to ensure high inter-rater reliability for scoring of the various behavioral events. Once trained, however, scoring is highly reliable and interrater reliability can be easily established using only a small subset of sample data.

Frame-by-frame video analysis, in contrast to traditional 3D motion tracking systems, can provide a more ethologically valid description of natural reaching and grasping behavior as it does not require the placement of markers or sensors onto the participant's arms or hands. Additionally, many 3D motion tracking systems require a constant and direct line of sight between the camera and sensors/markers placed on the hands. To ensure this, most users of this technology ask participants to begin the reach-to-grasp movement with the hand shaped in an unnatural configuration with the index finger and thumb pinched together. They also instruct the participant to grasp the target object in a pre-defined way (usually a pincer grip) with a pre-defined orientation. These directives are required to ensure that the reach-to-grasp movement unfolds in a predictable and stereotypical manner as traditional recording systems can suffer significant data loss when the trajectory of the arm and configuration of the hand do not follow a predictable course that maintains the line of sight between the camera and sensors/markers. Nonetheless, imposing these constraints severely limits the ethological validity of the task and can even alter the organization of the movement; for example, it is not possible to observe the key behavioral event of 'collection' when the initial hand configuration is that of a pinch between the thumb and index finger[13,14]. These limitations are largely overcome when using frame-by-frame video analysis as variations in reach trajectory and hand configuration are much less likely to result in a complete loss of data in the video record so there is no need to impose these unnatural constraints on the reach-to-grasp movement.

Frame-by-frame video analysis also makes it possible to observe subtle modifications of the reach-to-grasp movement beyond what is generally possible with traditional 3D motion tracking systems, especially when the modification is not a specific prediction of the study. An example will illustrate: **Figure 5** (top) shows measures of peak aperture acquired from participants reaching-to-grasp three different sized objects either with vision or without vision. The results suggest that participants preemptively scale peak aperture to match the size of the target in the Vision condition, but not in the No Vision condition. In the No Vision condition participants use a consistent peak aperture despite reaching for targets of varying size. If one were to consider *only* the type of data available from a traditional 3D motion tracking system, similar to that shown in **Figure 5** (top left), there are two possible explanations for this discrepancy. First, it could be that in the No Vision condition participants shape the hand into a grasping posture that matches the "average" or "middle" size of the three possible targets. Alternatively, they may not form a grasping posture at all, but rather, they may form a slightly more open hand during transport towards the target, to increase the chances of making tactile contact with the target, that coincidentally matches the size of the "middle" target. To differentiate between these two possibilities, it is necessary to review the data from the frame-by-frame video analysis, a sample of which is given in **Figure 5** (top right), which clearly indicates that the participants are not shaping their hand into a grasping posture that matches the "middle" sized object in the No Vision condition; rather, they are forming an open but neutral hand shape that could serve to either locate the target through tactile feedback and/or to grasp the target. Thus, frame-by-frame video analysis can provide clarification when data from traditional 3D motion capture systems is ambiguous and can enable a more veritable interpretation of the results.

The use of frame-by-frame video analysis to study the reach-to-grasp movements of unsighted adults[6,8,9,10], human infants[11], non-human primates[12], and rodents[15] has already greatly amplified our understanding of the neurobehavioral control of prehension. Specifically, the results of these studies have consistently shown that in the early stages of prehensile development and evolution the *touch-then-grasp* strategy, in which the Reach and Grasp are temporally dissociated to capitalize on tactile cues, is preferred over the *preshaping* strategy in which the two movements are integrated into a single seamless act under visual guidance. These results provide substantial behavioral support for the Dual Visuomotor Channel theory and further suggest that the theory should be revised to account for the fact that separate reach and grasp movements likely originate under tactile control long before they are integrated together under visual guidance[1,2].

Disclosures

The authors have no competing financial interests to disclose.

Acknowledgements

References

1. Karl, J.M., Whishaw, I.Q. Different evolutionary origins for the Reach and the Grasp: an explanation for dual visuomotor channels in primate parietofrontal cortex. *Front Neurol.* **4** (208) (2013).
2. Whishaw, I.Q., Karl, J.M. The contribution of the reach and the grasp to shaping brain and behaviour. *Can J Exp Psychol.* **68** (4), 223-235 (2014).
3. Jeannerod, M. Intersegmental coordination during reaching at natural visual objects. In: *Attention and Performance IX.* Long, J., & Badeley, A. eds., Hillsdale: Lawrence Erlbaum Associates, 153-169, (1981).
4. Arbib, M.A. Perceptual structures and distributed motor control. *Handbook of Physiology.* Volume **2**. Brooks, V.B., ed., American Psychological Society, 1449-1480, (1981).
5. De Bruin, N., Sacrey, L.A., Brown, L.A., Doan, J., Whishaw, I.Q. Visual guidance for hand advance but not hand withdrawal in a reach-to-eat task in adult humans: reaching is a composite movement. *J Mot Behav.* **40** (4), 337-346 (2008).
6. Karl, J.M., Sacrey, L.A., Doan, J.B., Whishaw, I.Q. Hand shaping using hapsis resembles visually guided hand shaping. *Exp Brain Res.* **219** (1), 59-74 (2012).
7. Domellöff, E., Hopkins, B., Francis, B., Rönnqvist, L. Effects of finger markers on the kinematics of reaching movements in young children and adults. *J Appl Biomech.* **23** (4), 315-321, (2007).
8. Karl, J.M., Sacrey, L.A., Doan, J.B., Whishaw, I.Q. Oral hapsis guides accurate hand preshaping for grasping food targets in the mouth. *Exp Brain Res.* **221** (2), 223-240 (2012).
9. Karl, J.M., Schneider, L.R., Whishaw, I.Q. Nonvisual learning of intrinsic object properties in a reaching task dissociates grasp from reach. *Exp Brain Res.* **225** (4), 465-477 (2013).
10. Hall, L.A., Karl, J.M., Thomas, B.L., Whishaw, I.Q. Reach and Grasp reconfigurations reveal that proprioception assists reaching and hapsis assists grasping in peripheral vision. *Exp Brain Res.* **232** (9), 2807-2819 (2014).
11. Karl, J.M., Whishaw, I.Q. Haptic grasping configurations in early infancy reveal different developmental profiles for visual guidance of the Reach versus the Grasp. *Exp Brain Res.* **232** (9), 3301-3316 (2014).
12. Whishaw, I.Q., Karl, J.M., Humphrey, N.K. Dissociation of the Reach and the Grasp in the destriate (V1) monkey Helen: a new anatomy for the dual visuomotor channel theory of reaching. *Exp Brain Res.* **234** (8), 2351-62 (2016).
13. Timmann, D., Stelmach, G.E., Bloedel, J.R. Grasping component alterations and limb transport. *Exp Brain Res.* **108** (3), 486-492, (1996).
14. Saling, M., Mescheriakov, S., Molokanova, E., Stelmach, G.E., Berger, M. Grip reorganization during wrist transport: the influence of an altered aperture. *Exp Brain Res.* **108** (3), 493-500, (1996).
15. Whishaw, I.Q., Faraji, J., Kuntz, J., Mirza Ahga, B., Patel, M., Metz, G.A.S., *et al*. Organization of the reach and grasp in head-fixed vs freely-moving mice provides support for multiple motor channel theory of neocortical organization. *Exp Brain Res.* **235** (6), 1919-1932 (2017).