Authoring Illustrations of Human Movements by Iterative Physical Demonstration

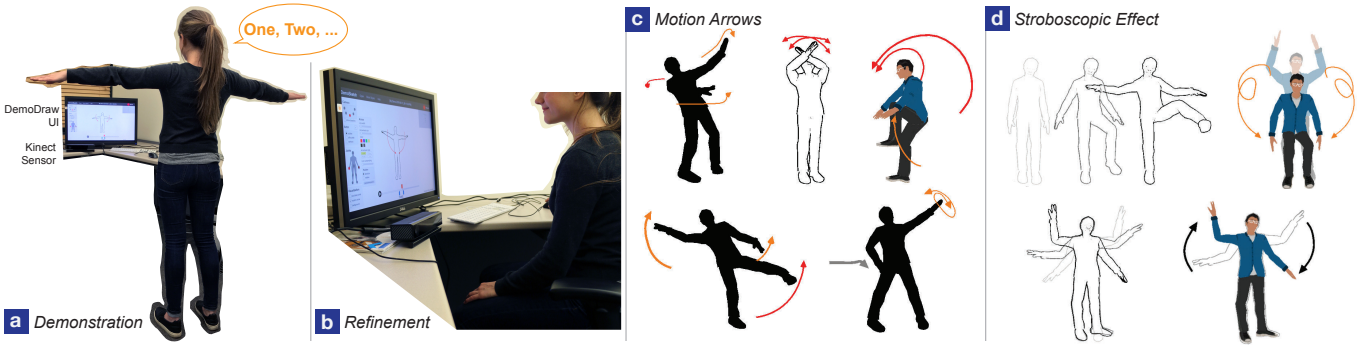

Figure 1. DemoDraw: (a) multi-modal "Demonstration Interface" to capture motion, verify results, and re-perform portions if needed; (b) conventional Refinement Interface for refinement and exploring other visualization styles; (c-d) examples of illustration styles.

ABSTRACT

Illustrations of human movements are used to communicate ideas and convey instructions in many domains, but creating them is time-consuming and requires skill. We introduce a multimodal approach for people to generate these illustrations by physically demonstrating the movements. Our DemoDraw system segments speech and 3D joint motion captured by a Kinect RGB-D sensor into a sequence of motion segments, each characterized by a key pose and salient joint trajectories. Based on this sequence, a series of illustrations is automatically generated using a stylistically rendered 3D avatar annotated with arrows to convey movements. During demonstration, users can also navigate existing illustrations using speech and amend or re-perform motions if needed. Once a suitable sequence of steps has been created, our system provides a Refinement Interface for finer control of visualization parameters. In a three-part evaluation, we validate the effectiveness of the generated illustrations and the usability of both the Demonstration Interface and the Refinement Interface. Our results show that our participants could create 4-7 step-by-step illustrations from demonstrations in 22 minutes on average.

Author Keywords

illustrations; instructions; tutorials; demonstrations; how-to; motion

INTRODUCTION

In sports, dance performance, and full-body or hand gesture-based interfaces, movement instructions are often conveyed with drawings of the body annotated with arrows or stroboscopic effects [18] (see Figure 6 for examples). These *illustrations of human movements* are also used within HCI to convey new user experiences in papers and storyboards [12]. When designed well, these illustrations can evoke a feeling of motion with reasonable precision, as long as the moving character is clearly represented with direction and magnitude [18].

We found that both professionals and non-designers create these kinds of illustrations, but the methods they use are commonly time-consuming and not amenable to iteration and editing. It can take between 10 minutes and several hours to construct scenes, pose and take photos, trace them, and adjust details like arrow placement. Moreover, identifying the appropriate pose and viewpoint ahead of time is hard. Unfortunately, changing these elements requires starting over again with new source photos.

Researchers have developed algorithms to visualize human motion [11, 15, 5, 6], but most prior work focuses on transforming datasets of pre-recorded motion capture sequences into visualizations for specific poses, not on authoring visualizations interactively. We extend an approach used in demonstration-based animation systems [8, 26, 24] where demonstration and authoring are integrated into one interactive workflow. These prior systems use proxy objects that a user manipulates to drive the animation of non-human objects and characters, usually with low degrees of freedom.

We propose DemoDraw, a system to enable people to rapidly create step-by-step motion illustrations through physical

demonstration. Our system uses iterative demonstrations (with additions, re-takes, and refinements) as a first-class interaction method. Authoring proceeds in two modes: *Demonstration*, performed using body motions and voice commands; and *Refinement*, which uses a standard desktop interface.

The user first records the motions to be illustrated by physically demonstrating them in front of a Kinect RGB-D sensor. As in current instructional practice, the user can simultaneously speak to denote important motion parts (e.g., "one, two, three, four"). The motions are then mapped to a 3D human avatar, which is shown with a concise look of an artist's line drawing of a body using Non-Photorealistic Rendering (NPR). We derived our rendering choices from a study of existing illustration practices. Speech and motion streams are analyzed to automatically segment motions into illustration figures with key frames. Salient joint movements are automatically identified and rendered as motion arrows overlaid on the body drawing (Figure 3b). During the *Demonstration* mode, segmented motions can be reviewed and re-recorded using speech commands. During *Refinement*, the annotation style and placement can be adjusted, camera angles moved, and alternate visualization styles explored in a mouse-driven GUI (see Figure 3c). A three-part evaluation with 14 participants shows that DemoDraw's illustrations are effective and people can use the Demonstration Interface and Refinement Interface to create illustrations of movements with various levels of complexity proficiently.

Generating illustrations of human movement instructions by in-situ demonstration, refinement, and editing has not been done before. Our work makes the following contributions:

- A novel approach for generating body motion illustrations through physical demonstration.
- Multi-modal interaction techniques to record, review, retake, and refine demonstration sequences.
- A set of methods for automatically analyzing 3D motion data with speech input to generate step-by-step illustrations.

RELATED WORK

Our work is related to research in demonstration-based authoring and motion visualization techniques.

Demonstration-Based Authoring

User demonstrations have been harnessed to generate explanatory, educational or entertainment media in domains including software tutorials [9, 22], animation [8, 26], 3D modeling [35], or physical therapy [34]. Systems can support different levels of integration between demonstration and authoring. Some focus on post-processing previously captured demonstrations, leaving no option to re-perform or refine demonstrations while authoring. Work that falls into this category includes: generating step-by-step software tutorials from video or screen recordings with DocWizards [9], Grabler et al.'s system [22], and MixT [13], and automatically editing and annotating existing video tutorials with DemoCut [14]. This "first demonstrate, then author" workflow is similar to graphics research transforming existing artifacts into illustrations. For example, using technical diagrams to generate exploded views or mechanical motion illustrations with systems by Li et al. [28] and Mitra et al. [32], using short videos to generate storyboards [20], creating assembly instructions by tracking 3D movements of blocks in DuploTrack [25] and closely related to our work, using existing datasets of pre-recorded motion capture sequences to generate human motion visualizations with systems by Assa et al. [5, 6], Choi et al. [15], and Bouvier-Zappa et al. [11].

Other systems integrate demonstration and authoring into one interactive workflow. This way iterative demonstrations (with additions, re-takes, and refinements) are a first-class interaction method for authoring. This general strategy was used by GENESYS [7], one of the earliest interactive computer animation systems. Authoring animation by demonstration remains a common approach, often using physical props as in Video Puppetry [8], 3D Puppetry [26], and MotionMontage [24].

While the primary goal of performance-based animation systems is to accurately track and re-target prop motions to virtual characters, DemoDraw focuses on the mapping from recorded body movement demonstrations to static illustrations conveying those motions. Some previous systems have also mapped body movement to static media: BodyAvatar [35] treats the body as a proxy and reference frame for "first-person" body gestures to shape a 3D avatar model; a Manga comic maker [29] maps the body pose directly into a comic panel. Systems that provide interactive guidance for teaching body motions could be considered the inverse of DemoDraw. Examples include YouMove [3] that teaches moves like dance and yoga, and Physio@Home [33] that guides therapeutic exercises.

Motion Visualization

Several of the systems above focus on developing automated algorithms to visualize various types of dynamic behavior, such as mechanical motion [28, 32], motion in film [20], and human movements [5, 11, 15]. Much of this work is inspired by formalizing techniques and principles for hand-crafted illustrations [2]. Similarly, we explore such principles for body movement diagrams in the following section.

The illustrations produced by our system are similar to those described by Bouvier-Zappa et al. [11], but our goals are different. Their goal is to automatically visualize large collections of pre-recorded motion capture sequences. We support many of the same visualization techniques in our system, including motion arrows, overlaid ghosted views, and sequences of poses, but our goal is to provide an interactive system that helps authors create illustrations for particular motions they wish to share with others. Since such demonstrations often involve mistakes and multiple repeated takes of the motion, DemoDraw supports interactions to help authors review and retake portions of their demonstrations for creating an illustration. Moreover, the interactive nature of DemoDraw enables finer-grained controls for adjusting visualization parameters and compensating for idiosyncratic characteristics of automated algorithms.

MOTION ILLUSTRATION PRINCIPLES AND METHODS

To understand motion illustration design and production, we surveyed related literature, studied found examples, and interviewed individuals who create such illustrations.

Design Principles

Cutting [18] argues that superimposing vector-like lines on an image satisfies four important criteria: it evokes a feeling of motion, the object undergoing motion is clearly represented, the direction of motion is clear, and the magnitude of motion is conveyed with reasonable precision. To complement this metaphoric representation, Cutting also argues for the more literal method of multiple stroboscopic images, which satisfies all criteria except clear motion direction. McCloud [30] provides further arguments and examples for using these methods in the field of comic illustration, and notes communication benefits when they are combined.

To examine how professional illustrators use motion lines and stroboscopic images, we gathered examples from sources like user manuals, gesture-based games, safety guides, illustration compendia (e.g., [31]) and how-to books (e.g., [23]). We found Cutting's notion of vector-like lines are almost always rendered with an arrowhead in a variety of styles (heads, weights, colors) with strokes typically two-dimensional, smooth, and offset to avoid occluding the object. Stroboscopic images can be overlapping or spatially distributed, and change in transparency or shading to convey time. The most common style for depicting the object undergoing motion is a simplified black-and-white contour drawing, but filled silhouettes and flat-shaded colour can also be found – using full color photographic detail is rare. By carefully removing extraneous details, such techniques help readers focus on only the salient, abstract information.

Interviews: Methods Used In the HCI Community

To understand current creation methods, we conducted video interviews with six Human-Computer Interaction researchers with experience creating motion illustrations. Conveying movement for interaction is common in HCI publications. We found 100 motion illustrations in 58 recent papers.

Findings. All interviewees used a similar methodology to create motion illustrations: they took still photographs of people performing actions, traced outlines using Adobe Photoshop (4/6) or Illustrator (2/6), then added graphic annotations to convey motion. All mentioned that it was time-consuming to set up scenes and poses, take and trace photos, then add details like arrow placement while maintaining a consistent style. Typical creation times were estimated between 10 minutes to a few hours. They also noted how difficult it was to make adjustments: changing the pose or viewpoint essentially meant starting over again with new source photos and re-tracing. Yet, identifying the best pose and viewpoint ahead of time is difficult and it often took several iterations to yield an illustration suitable for publication.

Design Space Goals and Workflow

Based on the above, we derive a canonical workflow to motivate the central design goal for our system. Authors face two primary illustration tasks (Figure 2): *defining the context* for portraying motion like the view of the body and salient aspects of motion; and *exploring a style of motion depiction* by choosing styles like lines-and-arrows or stroboscopic, then

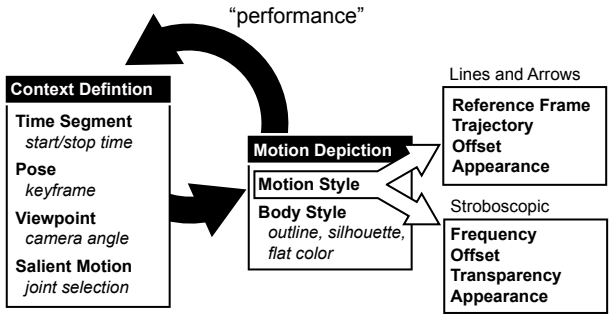

Figure 2. Canonical authoring workflow consisting of a Context Definition task then a Motion Depiction task. Design decisions associated with a task shown in bold with design parameters in italics.

adjusting related style parameters. These tasks and the underlying design parameters are highly interdependent, so authoring motion illustrations is necessarily an iterative process. This means that changes to one task parameter often leads to re-evaluating and changing the other. The problem with current methods, is that context is mostly "performed" using a time-consuming process of taking photos and manually tracing them. Therefore, the central design goal of our system is to make context definition low effort and iterative by using interactive demonstrations for automated context definition.

Designing a system to capture interactive demonstrations of *any* body movement also poses an input challenge. Since body movements form the demonstration itself, also issuing application commands with a body gesture introduces ambiguity. Using a hand held device, touch screen, or any conventional input is not ideal since performing requires open space and full freedom of movement. For these reasons, we use a multimodal voice and gesture interaction style traced back to Bolt's Put-That-There [10]. Like Bolt, we use voice for commands like *"start"* and *"stop"* with body movements providing command parameters in the form of the recorded demonstration, and for setting parameter context with utterances like *"one, two, three, four"* to label step-by-step segments.

DEMODRAW

DemoDraw enables authors to generate concise illustrations using two modes: the Demonstration Interface and the Refinement Interface. To provide an overview of how the system works, we present a scenario in which a motion illustration author, Marie, creates instructions for an 8-step dance tutorial.

In her living room, Marie begins using DemoDraw with the Demonstration Interface shown on her television by standing in front of a Kinect. In the center of the display, an avatar follows her movements in real-time (Figure 3a). This avatar is shown as an "outline" figure, but she could always change to different rendering effects like "silhouette" or "cartoon," or select a different 3D human model later using our Refinement Interface (Figure 3c).

Recording. Marie starts recording her physical demonstration via a voice command *"Record"* or *"Start."* After a 3-second countdown, DemoDraw captures the position, orientation, and

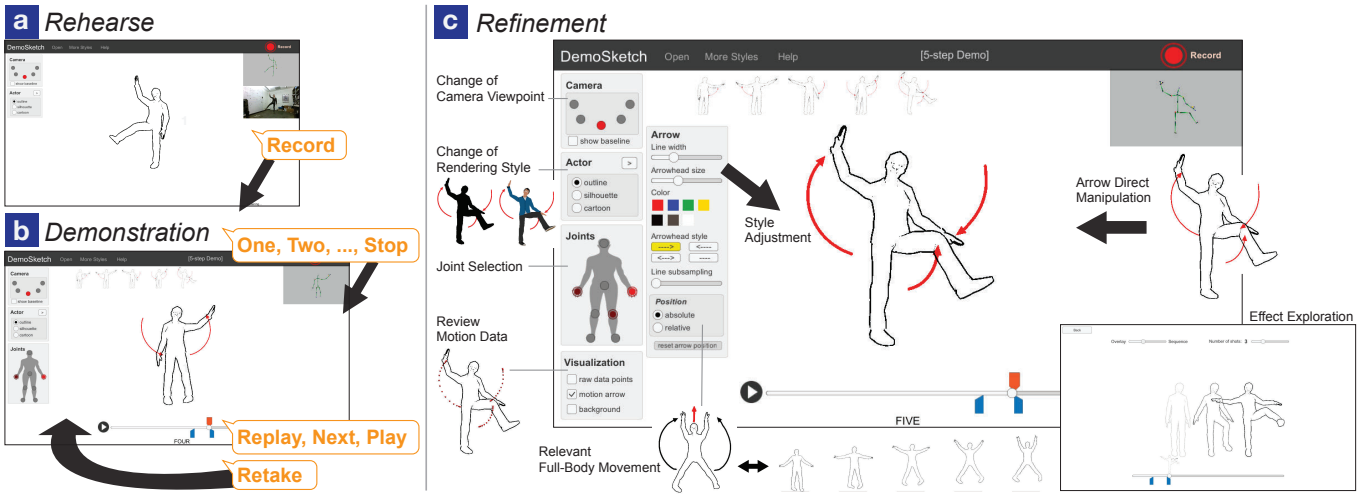

Figure 3. DemoDraw authoring UI: Using the Demonstration Interface, an author sees an avatar following her real-time movement (a). During a recording (initiated by voice command "Record", real-time feedback shows the speech labels. Once a recording is completed by voice command "Stop", the motion visualization and a timeline are immediately available (b) for the author to review, and a step-by-step overview will be generated. Later using the Demonstration Interface, the author can refine the visuals and explore more illustration effects (c).

depth distance of her body (using Kinect's simplified 25 body joint model). While demonstrating dance moves, Marie verbally indicates the count of each step with *"one, two, three, and four,"* just like she does when teaching a dance. The specific utterance is not constrained, Marie could use words like *"right, left, shake, and clap."* A speech recognition engine captures these labels with timestamps and displays them in the interface. Marie finishes recording by saying *"Stop"*. When a speech label is detected, DemoDraw automatically segments the motion around the utterance and identifies salient joints.

Reviewing and Re-Recording. After recording, an illustration of the first step of Marie's demonstration is rendered with motion arrows showing the path of the most salient joints. Figure 3b presents an example illustration that shows how her right hand waves from bottom to the top, and the left on the opposite direction. She also notices three panels emerged: A timeline below shows the start, end, and key frame points used to generate the current illustration, a side panel shows the visualized joints; an step-by-step overview of step snapshots is created and added to a motion sequence list. Marie can navigate to other illustrated steps by either saying *"Next"* or *"Back"*, or repeating one of the words she said during recording (like *"three"*) to skip to that corresponding step. To play an animation showing her continuous motion, she can say *"Play"* to play the current step only, or *"Replay"* to play the entire motion recording with each step visualization highlighted.

Once Marie reviews the steps, she realizes she should have exaggerated the hand motion in step 4. By saying *"Retake Four,"* Marie can re-record a partial sequence of movements including that step (e.g., redoing and saying "Four" and "Five"). When she ends the re-recording with *"Stop"*, the old illustration for that step is replaced with a new one (step four in this example) generated using the new motion recording.

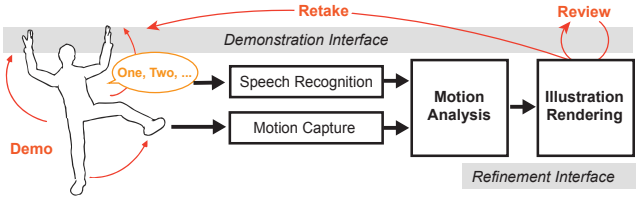

Figure 4. DemoDraw System Components and Pipeline

Motion Depiction Adjustments. Once Marie is satisfied with her demonstration, she walks out of the capture area to her desktop computer. The system automatically switches to Refinement Interface by revealing post-processing panels for detailed design refinements with a standard graphical user interface (Figure 3c). Using this interface, Marie adjusts several design parameters: the arrow appearance can be refined, including line width, arrowhead size, and color; the arrow offset can adjusted with direct manipulation dragging; the camera viewpoint can be adjusted by orbiting the camera to a side or three-quarter view; the joints used for motion paths can be added or removed using a panel; and the smoothed motion trajectory can be toggled on and off. She could also select a different key pose and adjust the start and end times of a motion segment by dragging the markers on the timeline. In addition, Marie could explore more illustration styles using stroboscopic rendering by selecting numbers of intermediate frames and how they render in one diagram. These results can be exported to step-by-step diagrams for a sharing purpose.

SYSTEM COMPONENTS AND IMPLEMENTATION

DemoDraw has four main components (Figure 4): a *motion capture* engine to record joint data from the author's demonstration and map it data to a 3D avatar; a *speech recognition* engine to process speech input for commands and motion

labels; a *motion analysis* algorithm that segments recorded motion and identifies salient joint movements for each illustration panel; and an *illustration rendering* engine to visualize the avatar and motion segments with different effects. The combination of these components enables an interactive and iterative system pipeline to translate demonstrations into motion diagrams. A notable technical contribution is our motion segmentation algorithm leveraging parallel speech label and joint motion input streams.

DemoDraw is implemented using C# in Unity 5[1]. It runs interactively on a Macbook Pro with Windows Bootcamp (2.5 GHz Intel Core i7 processor and 16 GB memory). Below we describe the design and implementation of each component.

Motion Capture

In support of our design goal to enable low-effort iteration within tasks, the motion capture component provides real-time feedback during demonstrations so authors can monitor their performance accordingly. We capture position and joint angles of a simplified 25-joint skeleton using a Kinect2 sensor and the Kinect SDK 2.0[2]. The real-time joint data is applied to a generic 3D human model (an "avatar") using forward kinematics enabled by a modified Unity asset[3].

Speech Recognition

Speech is used when recording a demonstration to label motions (e.g., "one, two, ...") and for recording and navigation commands (e.g. "start, stop" or "three, back" to skip to a motion segment) – see Figure 3 for the speech commands that DemoDraw supports. We recognize both types of speech using the Microsoft speech recognition library[4] to process audio captured by the Kinect microphone array. During recording, the time stamp, delay, and confidence of motion labels are logged for use in the the motion analysis algorithm.

Motion Analysis

Our motion analysis algorithm translates a multi-part demonstration recording into a sequence of labeled time segments, each with one or more salient joint motions and a keyframe of joint positions to use as a representative body pose (see Figure 5 for an illustration of the approach). Formally, the algorithm associates each speech label with a motion start and end time [T_s, T_e], set of salient joints $J_0, ..., J_n$, and keyframe time T_k. Each segment is sent to the Illustration Rendering engine to create each motion illustration in a multi-part sequence.

Human motion segmentation and activity understanding has been well studied in computer vision and graphics [1]. We adopted a time-space approach to identify salient motion sequences in 3D space. In our scenario, each movement may not be necessarily seen or labeled with a semantic meaning (such as "running" or "walking" in previous research). Therefore, our approach combines speech labeling from a user, similar to the scene segmentation method used by [14]. We make three assumptions about the synchronized data streams of speech labels and joint movements: authors make short pauses between motions to be grouped; the speech label utterances overlap or closely occur with at least one joint motion; step-by-step movements are clearly segmented without an overlap. These assumptions are practical as authors often pause for a moment to prepare for demonstrating the next movement in a step-by-step sequence.

Motion Segmentation. Let T_w be the delay-corrected time when a speech label like *"one"* was detected. The first operation is to determine [T_s, T_e] for each speech label such that $T_s - \varepsilon \leq T_w \leq T_e + \varepsilon$, where ε is set to be 1 second to allow short pauses between speech and movement. The operation begins by identifying periods of significant joint movement for 8 joints: the 5 end-effectors (head, hands, feet), 2 knees, and the body root. To filter jittery movements, joints are considered moving if smoothed inter-frame differences in absolute Euclidean distance are greater than a threshold. Specifically, for each joint J_i at time t, the average difference in position between two adjacent frames $\Delta P = |P_t - P_{t-1}|$ is computed over the subsequent half second (15 frames). If this moving average is greater than $0.05 meter/s$, then joint J_i is labeled as "moving" at time t, marked as m_i^t. This is repeated on all frames and all joints to find all periods of significant joint movement (pink rectangles in Figure 5).

Next, after combining all the consecutive $m_i^{T_s}, m_i^{T_s+1}, ..., m_i^{T_e}$ between time [T_s, T_e] for joint J_i, we begin labeling each joint moving period M_i: Given a list of speech labels, for each speech label at time T_w, label any unmapped period M_i of joint J_i where $T_s - \varepsilon \leq T_w \leq T_e + \varepsilon$ (i.e., the speech utterance is detected during or closely occur with a joint movement, illustrated as dashed lines crossing pink rectangles in Figure 5). Once all periods of joint movement are mapped to labels for all major joints, the motion segment start time and end time $[T_s, T_e]$ are set to the minimum start time and maximum end time across all mapped joint movement periods.

Joint Salience Identification. The salient joints for a segment is the set of all joints J_i that were mapped based on significant motion segments.

Key Pose Selection. We use a time near the end of each segment as the representative pose keyframe, specifically $T_k = T_e - 1$ second. We experimented with other options including the middle of the segment, but found designers most often chose a pose near the end as the key frame.

When retaking a partial demonstration with one or more speech labels, the full motion analysis algorithm is run on the new recording, and new motion segments are inserted into the sequence of labeled time segments.

Illustration Rendering

The Illustration Rendering engine generates a motion illustration for each segment of joint motion (bounded by [T_s, T_e]). There are two related rendering tasks: the body pose and the motion depiction style.

Body Pose. Following the principle of clarity through simplicity, Non-Photorealistic Rendering [21] algorithms are used

[1] `https://unity3d.com`

[2] `https://dev.windows.com/en-us/kinect`

[3] `https://www.assetstore.unity3d.com/en/#!/content/18708`

[4] `https://msdn.microsoft.com/en-us/library/hh361572`

to render the 3D human model as an outline, silhouette, or flat-shaded colour (see Figure 3c left for examples). The body pose is determined by all joint positions at keyframe time T_k.

Line and Arrow Depiction Style. Based on Cutting's criteria [18] and our survey of motion illustrations, we use lines with arrowheads as the default depiction style for visualizing joint movements. This style is rendered as follows: For each salient joint J_i in $J_0, ..., J_n$ of a motion segment, the absolute joint positions in world space over the period $[T_s, T_e]$ are used to construct a 3D poly-line. The default is to render a smooth poly-line by interpolating positions using Catmull-Rom. Two 3D cones are positioned collinear with the last two polyline positions to form arrowheads for both the beginning and the end of a line. Although the poly-line is constructed in 3D, it is rendered with shading to appear 2D.

For some motions, visualizing absolute joint positions might not be suitable. For example, for a two-foot jump with a two-hand waving motion (see Figure 3c bottom), our algorithm will mark all major joints as salient and generate multiple arrows showing the jump movement, but failing to convey the hand waving. Authors can choose to visualize joint motions *relative* to the spine instead. The same motion analysis algorithm described above can then be re-run using relative motion. For the two-foot jump example, this would render a more concise illustration with a single up arrow (for the overall jump direction) and two curve arrows (for the hand movements).

Authors can review the results using both the Demonstration Interface and Refinement Interface. With the latter, line weight, arrowhead sizes, and color can also be adjusted and re-rendered in real-time using a graphical interface (see Figure 3c left). Arrows can also be re-positioned by direct manipulation dragging.

Stroboscopic Depiction Style. Cutting [18] noted stroboscopic effects are also effective, and we found examples of illustrations with a sequence of overlaid semi-transparent body poses in our survey. Therefore, authors can select a stroboscopic depiction style in the Refinement Interface. The style is rendered by compositing multiple semi-transparent renderings of intermediate body poses between T_s to T_e behind a rendering of the representative pose at keyframe time T_k. Authors can adjust the number of intermediate poses (the default is 3 poses) and the horizontal offset between each intermediate pose rendering can be adjusted to stack them up (zero offset by default) or spread them out.

Results

The DemoDraw pipeline is capable of generating expressive and clear motion illustrations. In Figure 1c, motion arrows show the upper body motion (top left), hand waving back and forth (top middle), and hand circular motion (bottom right). Whole body motions can also be visualized (bottom left), and can be especially helpful when motions are best viewed from a different angle, such as the side view (top right). In Figure 1d, stroboscopic effect depicts the transition from the start pose to the end pose, which can be rendered as a sequence (top left) or in one combined pose (bottom left). A combination of this effect with motion arrows enhances an illustration efficiently in an integrated figure (top and bottom right).

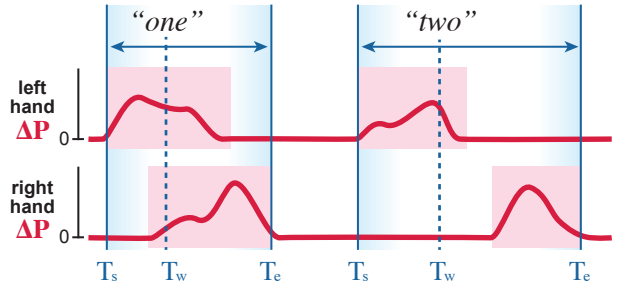

Figure 5. Illustration of motion analysis algorithm (two joints shown due to space): significant periods of joint motion (pink) are mapped to speech labels to define motion segments (blue). Note the right hand period is mapped to *"two"* because it begins shortly after the left hand period.

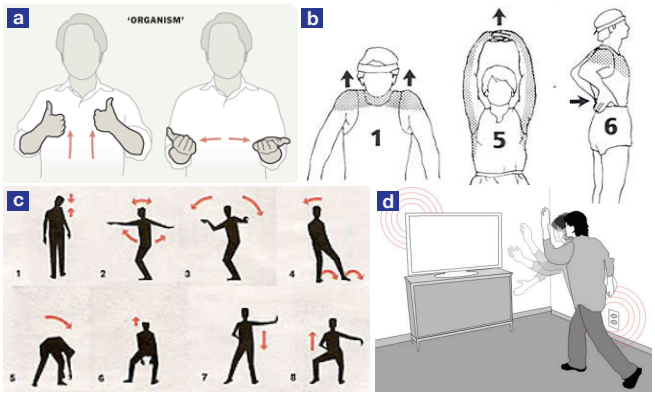

Figure 6. Examples of manually generated movement diagrams from print and online materials for general audiences (a for sign language [17], b for weight training [4], c for Michael Jackson's Thriller dance moves [author unknown]); and from HCI publication (d for gestural interfaces [16]).

USER EVALUATION

To evaluate the capability and usability of DemoDraw, we conducted three user studies. Since our motivation is to create illustrations that can be understood and learned, the first study with 10 participants tested the effectiveness of motion illustrations generated by DemoDraw. The second study with the same 10 participants evaluated the Demonstration Interface for recording and refining motion demonstrations. The third study with 4 different participants evaluated the Refinement Interface for refining and editing a recording that was already captured. We describe the details of each study with results in the sections below.

Study 1: Illustration Effectiveness

Hypothesis. Learners with no prior experience can understand and re-perform motions after reviewing step-by-step illustrations generated by DemoDraw.

Participants. We recruited 10 participants (5 females), aged 18 to 33 years (M=24.3), from a university and an IT company. Six participants had previously created illustrations (from 5 to

		Intended motion	Performed motion	# of Users	Participant Explanation
Set 1 Step 2		Waving hands back and forth	Waving one way (crossing hands)	2 / 10	Didn't catch it
Set 1 Step 3		Hands circling out and in	Hands circling in and then out (opposite direction)	4 / 10	Didn't catch it
		Squatting	N/A (missed)	7 / 10	(6/7) focused on hand motions and did not see the down arrow; (1/7) noted the arrow but thought it referred to the hands
Set 2 Step 1		Moving right hand from lower left to upper right	Moving right hand from lower right to upper right	5 / 10	(4/5) Didn't notice the start position; (1/5) assumed the starting pose is a stand position

Table 1. Incorrect movements performed by participants in Study 1.

50 diagrams), but none involved body motion. The common creation tool they used was Adobe Illustrator. Among all participants, four were comfortable performing dance moves.

Experimental Setup. The study was conducted in a lab environment with static indoor lighting. A video camcorder was set up to record participants.

Procedure and Tasks. To help introduce participants to the context of our work, we first provided existing examples of movement illustrations shown in Figure 6. Then, we presented two sets of printed motion diagrams generated by the experimenters using DemoDraw. Each set included 4 steps (see A.1). For each set, we asked participants to interpret the illustrations and perform the depicted sequence in front of the camcorder when they were ready. On average, the study lasted 4 minutes; participants practiced one minute and half minute respectively for the two sets prior to recording.

Measures. We coded the video recordings as follows: for each joint movement in a step, we measured if 1) user intentionally moved the joint, i.e., a hit or a miss, 2) the start and end positions approximately matched the shown positions (e.g., moving right hand from lower left to upper right to the body), and 3) the movement from start to end positions was performed in the same way as shown (e.g., moving hand straight). We also marked if users intentionally moved other joints in addition to highlighted ones. We did not code the speed of each motion as our illustrations did not convey this element.

Study 1 Results

All participants successfully performed at least 5 of the 8 steps, with 10 of the 16 joint movements being correct (62.5%). Table 1 lists systematic errors made by our users and explanations they provided for them. Bi-directional arrows, motions involving independent trajectories of multiple body parts, following complex trajectories, and inferring starting positions from arrows were all sources of errors. We suspect that some of the misinterpretations from our illustrations could be clarified via other visualization styles that DemoDraw is able to generate. We validate this in Study 3.

Study 2: Demonstration Interface

Hypothesis. Amateurs can efficiently create step-by-step motion illustrations using DemoDraw's multi-modal Demonstration Interface.

Participants. The 10 participants from Study 1 completed this study immediately after the prior study. Completing both study 1 and 2 took 45-60 minutes per participant.

Experimental Setup. The study was conducted in the same space as Study 1. The DemoDraw system ran on a Windows 8.1 machine, connected to a 30-inch monitor and external mouse and keyboard. The Kinect sensor was placed 3-feet above the floor to capture a 8×8-feet clean office space.

Procedure and Tasks. The study began with a learning phase where participants were introduced to DemoDraw and shown how to create the second set of illustrations from Study 1 by the experimenter. Participants were then asked to record the same motions and review the results. This learning phase lasted 5-10 minutes.

Then participants completed an evaluation phase with four tasks in sequence:

1. The experimenter physically demonstrated a set of 4 moves for a gestural interface with the right hand, including waving, circling, making a reversed V shape, and swiping to the right (see A.2-1). Participants were asked to record and review the captured results. Once they were satisfied with the recording, they were asked to rate each illustration step automatically generated by our syste,
2. Similar to task 1, but with a set of 8 dance moves (see A.2-2).
3. The experimenter introduced the retaking operation and asked participants to choose one step from task 2 that they would like to revise. Participants re-performed part of the motion and reviewed.
4. Participants were asked to perform any 4 to 8 moves they could imagine and retake them until they were satisfied with the results (within a time limit of 5 minutes).

For each task, the numbers of attempts were not restricted.

Measures. In tasks 1 and 2, participants rated each step along five dimensions: *"Q1: The visualization accurately captured/described my motion"* (a 5-point Likert scale from "1: Strongly disagree" to "5:Strongly agree"), *"Q2: The visualization shows all the important joints of movement"*, *"Q3: It shows at least one extraneous joint"*, *"Q4: The key pose was appropriately chosen"* (Q2, Q3, and Q4 were answered "Yes", "No", or "N/A"), and *"Q5: This figure needs more (manual) editing before I would share it with others"* (choose from "1: Definitely needs edits" to "5: Very comfortable to share as is").

An online questionnaire was provided at the end of the study. Answers to both Likert-scale questions and comments were collected.

Study 2 Results

On average, participants completed task 1 in 5 minutes with $\mu = 2.3$ takes, task 2 in 10 minutes ($\mu = 2.8$ takes), task 3 in 3 minutes ($\mu = 2.3$ takes), and task 4 in 5 min ($\mu = 2$ takes). A.2 and A.3 provide examples of illustrations created by participants. Below we discuss participant feedback on the illustrations and the system design.

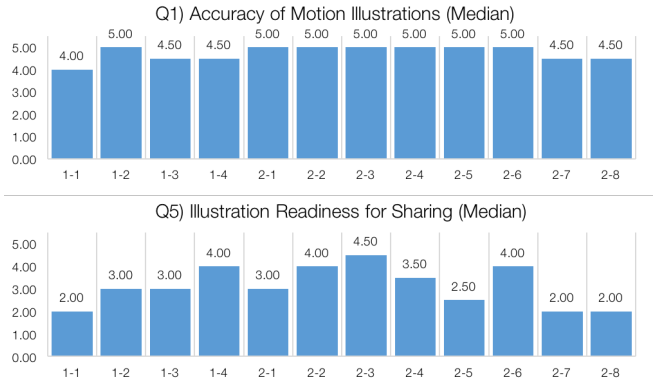

Figure 7. User feedback from Study 2, showing the median rating of each step illustration.

Ratings and Accuracy. Overall, participants thought the illustrations accurately described the motions (average median rating of 4.5 and 4.88 for task 1 and 2 respectively from Q1, where 5 = strongly agree), but on average, 4 of the 12 steps required further editing in order to share they illustrations with others (where medians were below 3 in Q5). Figure 7 shows the results of step ratings. P8 comments, *"This figure represented the overall motion well (...) In particular, it captured all key poses, and the motion lines are easy to follow."* But P1 mentioned, *"the system picked up really small movements in my other joints that were not relevant to the motion I was trying to depict (such as a small motion in my wrist or elbow)."*

In the ratings of a total of 120 steps from 10 participants, all but one visualization (99%) showed all the important joints (Q2), and in 80% of cases the illustrations precisely selected only the salient joints without extraneous movements (Q3). Participants explained: *"the picture correctly represents my stance and body position. the arrows are easy to see and follow"* (P1), *"the lines were very accurate"* (P2), and *"The arcs are gorgeous and represent the intention of my motion really well."* (P5)

Several participants mentioned that they appreciated how DemoDraw smoothed the arrows, especially when their demonstration was not perfect or there was some capture noise. Some noted the motion data capturing in 3D as an advantage. During the debrief session, some participants were shown the illustrations from different view points as their motion involved changes in depth. As P9 noted: *"I love the multiple camera angles for the wiggle arm motion I did in step 1."*

Key Pose Selection. Participants found 94% of key poses were selected correctly (Q4), commenting for example: *"The key poses are very descriptive of the motion"* (P2), and *"The key frames were just right"* (P5).

Authoring Workflow. Participants found it easy to learn DemoDraw (Median=5 out of 5) and easy to create illustrations using our system (Median=4.5). All participants were able to author and navigate via the speech interface. Specifically, P10 noted, *"The voice command allows people be able to control the system remotely. Without the voice capability, the system can be impractical in the single-person use case."*

Participants were especially impressed by how fast authoring could be to generate a step-by-step diagram: *"Surprisingly fast to make some really cool full-body motion demonstrations. There is no way I could do this in higher-fidelity than a napkin sketch in the same time"* (P5) and *"the system saves significant amount of time creating illustrations"* (P10). We also asked participants to estimate the time required if they were to generate a similar 8-step diagram without using DemoDraw, four participants answered that they would not be able to create manually, while others responded that it would take 90 to 160 minutes as each single figure might take 10-20 minutes.

Improving a Demonstration. The immediate visual feedback during the capturing phase was effective in helping authors review, adjust, and retake their performances. P4 explained, *"I learned how to exaggerate the important aspects of motion without being explicitly told to."* In task 3, when we introduced the retaking capability, participants explained that this function would be especially helpful for a long motion sequence. All but one participant chose to retake step 2-7, which involved a holding position with one foot. Three participants later used the same technique for task 4. P8 also noted that it was helpful as *"I could improve this by retaking that step and moving smoothly"* when referring to a specific pose that she thought needed additional work.

Study 3: Refinement Interface Effectiveness

To understand how users refine automatically-generated results and generate different visualization styles in DemoDraw's Refinement Interface, we conducted an informal study with 4 participants (all males, aged 23 to 32 years, M=28) from an IT company. The same experiment setup as Study 2 was used.

We provided a warm-up task where participants loaded one motion recording, which showed a continuous movement of lifting the right hand up and pushing to the right. Experimenters guided them to create an illustration with a thick arrow showing the hand movement using DemoDraw. Manipulations included the motion segmentation boundary, selecting joints of interests, line width and color, visual styles, and viewpoints. This learning phase was 5 minutes on average.

Four tasks were given in sequence: In tasks 1 and 2, participants were given a motion recording and three illustrations created using our system. We presented the printed figures one by one and asked them to approximate them with DemoDraw. In tasks 3 and 4, participants used DemoDraw to physically perform and record two specific motions that experimenters introduced. They then used our system to create an illustration that they thought would best convey the motion. They were free to use any skills they had learned from the previous tasks without time limitation. Finally, an online questionnaire was provided.

Study 3 Results

Unlike Study 2, which used only one illustration style with motion arrows, this study provided more versatile visualization techniques and manual refinement. We noted that participants actively tweaked visual parameters and the styles using the available options from the Refinement Interface, especially when creating the stroboscopic effect. Decisions

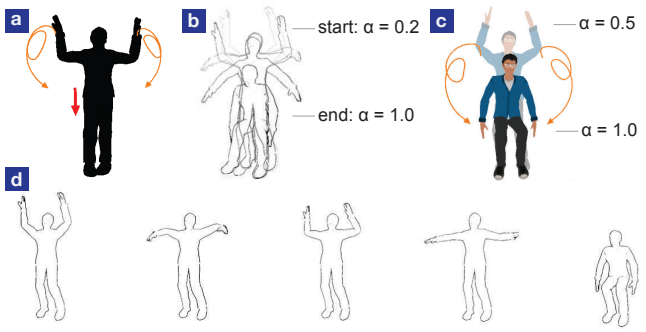

Figure 8. Different illustration effects conveying the same motion recording using DemoDraw's Refinement Interface: a and c are created by the paper authors and a was used in Study 1; b by Study 3-P1 using 4 intermediate frames with zero offset; d by Study 3-P2 using 5 frames, positioned as a sequence.

include: numbers of intermediate frames and the layout, dragging to reposition the arrows, and arrow color and width. In addition, participants had strong preferences on styles once they had several visualization options. For example, P4 explicitly commented on the stroboscopic effect of task 3 as *"this is exactly what I looked for – It clearly conveys the start and end poses."* P2 preferred the cartoon renderer than silhouette as *"this character looks just like me!"* They indicated that they were not able to create similar illustrations without using DemoDraw (Median=2). These results indicated that the detailed editing aspect of our system through the Refinement Interface was effective in authoring illustrations that could be expressive for various types of motions.

DISCUSSION

Participants were clearly excited about the overall experience using both the Demonstration Interface and the Refinement Interface. Some explicitly pointed out their enjoyment when using our system: *"it accurately captures how much fun I had making it. :)"* (Study 2-P9), and *"For professional artists, the system not only increases their productivity, but also brings joy and fun to this kind of tasks"* (Study 2-P10).

Participant feedback suggests motion illustrations generated by DemoDraw are expressive enough to depict their demonstrations. Our multi-modal interface with our motion analysis and rendering algorithms enabled users to quickly create step-by-step diagrams. Recall that current methods using existing software tools to create similar diagrams take significant time and making visual or spatial changes is difficult.

Illustration Styles

In Study 1, motion arrows successfully conveyed the majority of movements. For some movements when arrows failed to express the intent, other illustration effects might clarify the details to depict the start, intermediate, and end poses. For example, in the first motion set of Study 1, the circular hand movements and squatting action of Step 3 might not be easily interpreted (see Table 1), but for the same motion, participants preferred stroboscopic effect that clearly showed the arm movements and transition in height (see Figure 8).

Furthermore, as DemoDraw captures the continuous motion sequence in 3D from a demonstrator, our system also generates animations showing the dynamic movements. In the warm-up task of Study 2 that captured the second motion set in Study 1, some participants explained that the playback animation of the recording clarified the motion where they incorrectly interpreted the start position. We propose that as motion arrows can efficiently and effectively express most of the motions, a mixed-media version can be created that has been shown to be useful for clarifying step-by-step instructions [13], where viewers can selectively review part of a static diagram with in-place animation playback. In addition, the 3D reconstruction also makes it possible to review motions from different angles. All in all, our technology enables both instructors and viewers to interactively create and review motion illustrations in multiple ways.

LIMITATIONS AND FUTURE WORK

Our current system has several limitations based on both architectural decisions and limits in available technology.

Limited Interactions in Demonstration Mode. Presently, authors can review and retake steps using voice commands, but many fine-grained operations are only available in the Refinement Interface, which requires them to leave the performance area. Future work should investigate if voice commands combined with gestures can expose more functionality like timeline scrubbing to the author, to tighten the feedback look between performance, context setting, and depiction.

Motion Capture and Segmentation. The quality of DemoDraw illustrations is limited by the accuracy of motion capture data. Consumer devices such as the Kinect depth sensor we employed are widely available, but they suffer from problems with joint occlusions and fine-grained hand tracking. We found that illustrating such motions often required multiple recordings to obtain an artifact-free performance. Our segmentation algorithms currently assume that motions are separated by periods of inactivity, so we cannot yet capture and segment continuous motions that might be necessary, e.g., in different sports, where interruptions are not possible. Retargeting motion from a human performer to an avatar can also introduce artifacts when skeletal geometry does not match. Future work could apply retargeting approaches from the computer graphics literature [19] or examine if it is feasible to automatically generate suitable avatars that match performers' anatomy more closely.

Movements Involving Objects and Multiple Users. Many illustrations focus on motions while holding props (e.g., a tennis racket or baseball bat in sports) or the manipulation of objects (e.g., furniture assembly). We do not yet support such motions. One reason is that our implementation only tracks skeletons, not hands or their interactions with objects. While the general case seems very hard, using techniques for recognizing objects in video based on a library of 3D models [27] appears promising. Our current implementation is for single user, but we argue that it is possible to include multiple performers by loading and controlling additional avatar models, which would be especially useful in dancing.

Interpretability of Motions. DemoDraw can visualize the trajectories of multiple joints in a single image, but does not yet take the different timing of sub-motions into account. This can make illustrations of complex motions hard to interpret. Future work could provide per-joint timelines and automatically number sub-motions by their start times. In addition, the dynamics of motion are not adequately represented in output images. To address this, we have begun to experiment with mixed-media output formats. Inspired by MixT [13], we can render static illustrations that can replay a motion segment as an animation when clicked.

CONCLUSION

We introduced DemoDraw, a multimodal system for generating motion illustrations by physically demonstrating the movements. DemoDraw translates speech and 3D joint motion captured by a Kinect RGB-D sensor into a segmented sequence of key poses and salient joint movements. A step-by-step diagram showing a series of motion illustrations is automatically generated using a stylistically rendered 3D avatar annotated with arrows to convey movements. DemoDraw's multi-modal Demonstration Interface enables authors to record, review, and retake physical movements, and later refine and explore different motion visualizations with a Refinement Interface. We believe the Demonstrate-Refine distincion generalizes to other demonstration-based interfaces beyond motion illustrations. The primary motivation of this work is to provide users with domain-appropriate authoring tools that free them from tedious low-level tasks and allow them to focus their effort on both communicative and aesthetic aspects. We look forward to applying the same approach to other instructional materials and illustration types in the future.

REFERENCES

1. J.K. Aggarwal and M.S. Ryoo. 2011. Human Activity Analysis: A Review. *ACM Comput. Surv.* 43, 3, Article 16 (April 2011), 43 pages. DOI: http://dx.doi.org/10.1145/1922649.1922653
2. Maneesh Agrawala, Wilmot Li, and Floraine Berthouzoz. 2011. Design Principles for Visual Communication. *Commun. ACM* 54, 4 (April 2011), 60–69. DOI: http://dx.doi.org/10.1145/1924421.1924439
3. Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 311–320.
4. Robert Anderson and Jean Anderson. 2002. Before and After Weight Training. (2002).
5. Jackie Assa, Yaron Caspi, and Daniel Cohen-Or. 2005. Action synopsis: pose selection and illustration. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 667–676.
6. Jackie Assa, Daniel Cohen-Or, I-Cheng Yeh, Tong-Yee Lee, and others. 2008. Motion overview of human actions. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 115.
7. Ronald M. Baecker. 1969. Picture-driven animation. In *Proceedings of the May 14-16, 1969, spring joint computer conference*. ACM, Boston, Massachusetts, 273–288. DOI: http://dx.doi.org/10.1145/1476793.1476838
8. Connelly Barnes, David E. Jacobs, Jason Sanders, Dan B Goldman, Szymon Rusinkiewicz, Adam Finkelstein, and Maneesh Agrawala. 2008. Video Puppetry: A Performative Interface for Cutout Animation. In *ACM SIGGRAPH Asia 2008 Papers (SIGGRAPH Asia '08)*. ACM, New York, NY, USA, 124:1–124:9. DOI: http://dx.doi.org/10.1145/1457515.1409077
9. Lawrence Bergman, Vittorio Castelli, Tessa Lau, and Daniel Oblinger. 2005. DocWizards: A System for Authoring Follow-me Documentation Wizards. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA, 191–200. DOI: http://dx.doi.org/10.1145/1095034.1095067
10. Richard A. Bolt. 1980. Put-that-there: Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '80)*. ACM, New York, NY, USA, 262–270. DOI: http://dx.doi.org/10.1145/800250.807503
11. Simon Bouvier-Zappa, Victor Ostromoukhov, and Pierre Poulin. 2007. Motion cues for illustration of skeletal motion capture data. In *Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*. ACM, 133–140.
12. Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
13. Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. 2012. MixT: automatic generation of step-by-step mixed media tutorials. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 93–102.
14. Pei-Yu Chi, Joyce Liu, Jason Linder, Mira Dontcheva, Wilmot Li, and Bjoern Hartmann. 2013. Democut: generating concise instructional videos for physical demonstrations. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 141–150.
15. Myung Geol Choi, Kyungyong Yang, Takeo Igarashi, Jun Mitani, and Jehee Lee. 2012. Retrieval and visualization of human motion data via stick figures. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 2057–2065.
16. Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Humantenna: using the body as an antenna for real-time whole-body interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1901–1910.
17. Jonathan Corum. 2012. Drawing Science in Sign. (2012). http://style.org/sign/

18. James E. Cutting. 2002. Representing motion in a static image: constraints and parallels in art, science, and popular culture. *Perception* 31, 10 (2002), 1165–93. DOI: http://dx.doi.org/10.1068/p3318
19. Michael Gleicher. 1998. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM, 33–42.
20. Dan B Goldman, Brian Curless, David Salesin, and Steven M Seitz. 2006. Schematic storyboarding for video visualization and editing. In *ACM Transactions on Graphics (TOG)*, Vol. 25. ACM, 862–871.
21. Amy Gooch, Bruce Gooch, Peter Shirley, and Elaine Cohen. 1998. A non-photorealistic lighting model for automatic technical illustration. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*. ACM, 447–452.
22. Floraine Grabler, Maneesh Agrawala, Wilmot Li, Mira Dontcheva, and Takeo Igarashi. 2009. Generating photo manipulation tutorials by demonstration. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 66.
23. S. Greenberg, S. Carpendale, B. Buxton, and N. Marquardt. 2012. *Sketching User Experiences: The Workbook*. Elsevier/Morgan Kaufmann. https://books.google.com/books?id=c-RAUXk3gbkC
24. Ankit Gupta, Maneesh Agrawala, Brian Curless, and Michael Cohen. 2014. MotionMontage: A System to Annotate and Combine Motion Takes for 3D Animations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2017–2026. DOI: http://dx.doi.org/10.1145/2556288.2557218
25. Ankit Gupta, Dieter Fox, Brian Curless, and Michael Cohen. 2012. DuploTrack: A Realtime System for Authoring and Guiding Duplo Model Assembly. In *Proceedings of the 25th annual ACM symposium adjunct on User interface software and technology*. ACM, New York, NY, USA, 13.
26. Robert Held, Ankit Gupta, Brian Curless, and Maneesh Agrawala. 2012. 3D puppetry: a kinect-based interface for 3D animation.. In *UIST*. Citeseer, 423–434.
27. Natasha Kholgade, Tomas Simon, Alexei Efros, and Yaser Sheikh. 2014. 3D object manipulation in a single photograph using stock 3D models. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 127.
28. Wilmot Li, Maneesh Agrawala, Brian Curless, and David Salesin. 2008. Automated generation of interactive 3D exploded view diagrams. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 101.
29. David Lumb. 2013. "Manga Generator" Uses The Kinect To Put Your Smooth Moves In A Custom Comic. (Sept. 2013). http://www.fastcolabs.com/3016870
30. Scott McCloud. 1994. *Understanding Comics* (reprint edition ed.). Avon, New York.
31. P. Mijksenaar and P. Westendorp. 1999. *Open here: the art of instructional design*. Joost Elffers Books. https://books.google.com/books?id=fsJVAAAAMAAJ
32. Niloy J Mitra, Yong-Liang Yang, Dong-Ming Yan, Wilmot Li, and Maneesh Agrawala. 2010. Illustrating how mechanical assemblies work. *ACM Transactions on Graphics-TOG* 29, 4 (2010), 58.
33. Richard Tang, Hesam Alizadeh, Anthony Tang, Scott Bateman, and Joaquim A.P. Jorge. 2014. Physio@Home: Design Explorations to Support Movement Guidance. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, New York, NY, USA, 1651–1656. DOI: http://dx.doi.org/10.1145/2559206.2581197
34. Ross Yeager. 2013. *An Automated Physiotherapy Exercise Generator*. Master's thesis. EECS Department, University of California, Berkeley. http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-91.html
35. Yupeng Zhang, Teng Han, Zhimin Ren, Nobuyuki Umetani, Xin Tong, Yang Liu, Takaaki Shiratori, and Xiang Cao. 2013. BodyAvatar: Creating Freeform 3D Avatars Using First-person Body Gestures. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 387–396. DOI: http://dx.doi.org/10.1145/2501988.2502015

APPENDIX

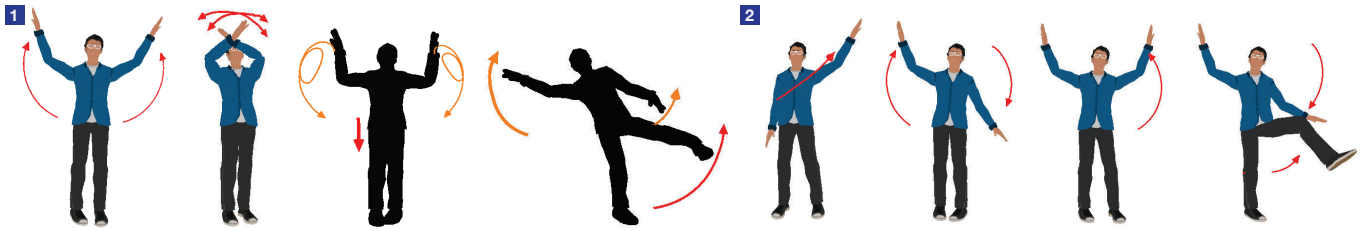

Figure A.1. Tasks provided in Study 1: We showed the printouts of these two sets of 4-step motions generated by DemoDraw using both the Demonstration Interface and the Refinement Interface. We asked participants to re-perform in front of a camera.

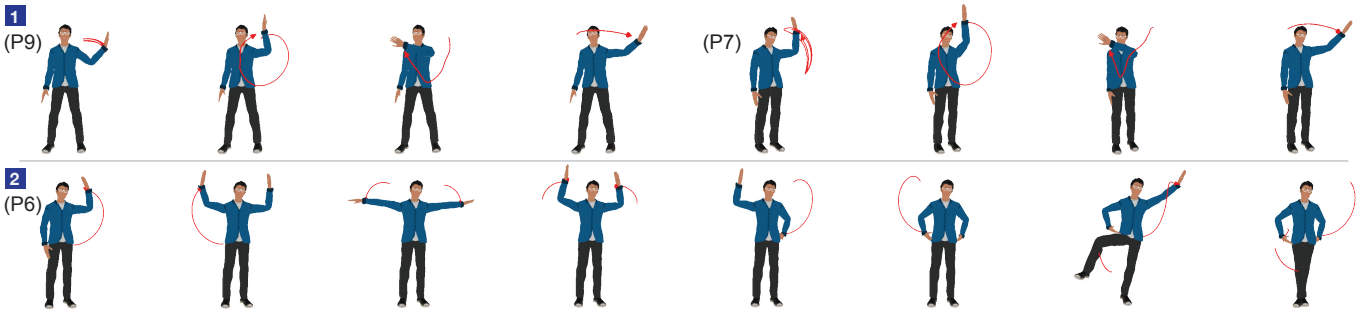

Figure A.2. Step-by-step illustrations generated by participants in Study 2 using the Demonstration Interface: 1) Results from P9 and P6 showing the same four gestures of a gestural interface in task 1, and 2) Results from P6 showing 8-step moves in task 2.

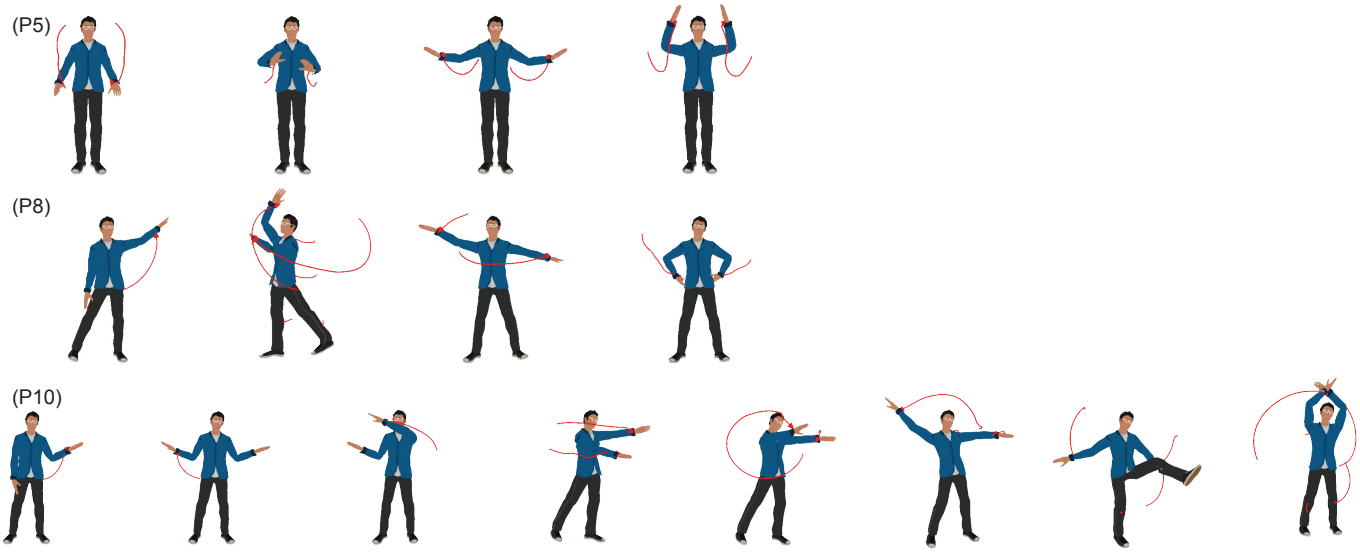

Figure A.3. Selected illustrations from the open-ended task created by three different participants using the using Demonstration Interface in Study 2: P5 performed to conduct a 4/4 beat pattern; P8 and P10 each performed four and eight free moves.