Exp Brain Res (2008) 187:139–152
DOI 10.1007/s00221-008-1291-2

RESEARCH ARTICLE

A new way to quantify the fidelity of imitation: preliminary results with gesture sequences

Brian J. Gold · Marc Pomplun · Nichola J. Rice · Robert Sekuler

Received: 10 July 2007 / Accepted: 15 January 2008 / Published online: 15 February 2008

Abstract Imitation is a common and effective way for humans to learn new behaviors. Until now, the study of imitation has been hampered by the challenge of measuring how well an attempted imitation corresponds to its stimulus model. We describe a new method for quantifying the fidelity with which observers imitate complex series of gestures. Wearing a data glove that transduced movements of their digits, subjects viewed and then reproduced a sequence of gestures from memory. The velocity profile of each digit's flexion or extension was used to segment movements made during an imitation into gestures that can be compared against corresponding gestures in the stimulus model. The outcome is a multivariate description of each imitation, including its temporal characteristics, as well as spatial errors (in individual gestures and in the ordering of those gestures). As a demonstration, we applied this method to data from an imitation learning experiment with gesture sequences. With repetition, overall fidelity of imitation improved, with various aspects of the imitation improving at different rates. Confirming the approach's usefulness, when we varied the complexity associated with imitation, that variation was robustly reflected in our measures of imitation quality. Finally, we describe a simple way to extend our methods to make them useful not only in assessing imitation and imitation learning, but also in various settings in which the detection and characterization of subtle abnormalities in movement production is paramount.

Keywords Imitation · Praxis · Sequence learning · Motion capture · Motor learning

Supported by NSF grant SBE-0354378. A preliminary version of this work was presented in 2007 at the Eleventh International Conference on Cognitive and Neural Systems, Boston, MA.

B. J. Gold · N. J. Rice · R. Sekuler (✉)
Volen Center for Complex Systems, Brandeis University,
415 South St, MS 013, Waltham, MA 02454, USA
e-mail: sekuler@brandeis.edu

M. Pomplun
Department of Computer Science,
University of Massachusetts at Boston,
100 Morrissey Blvd, Boston, MA 02125, USA

Introduction

By observing and then attempting to imitate the actions of others, humans acquire many cognitive competencies, including language and various skilled behaviors. After a long period of relative dormancy, research interest in imitation has been rekindled by the discovery of "mirror" neurons in the macaque brain (Rizzolatti et al. 1996). A putative analogue in the human brain, the "mirror system," has been assigned various functions (Iacoboni 2005), including central roles in action imitation, rehearsal (motor imagery) of actions, as well as in understanding the intentions of others (but see, Oztop et al. 2006). However, despite its importance for neuroscience and related disciplines, imitation has long resisted the kind of quantitative study that would be required for full understanding of the neural and cognitive mechanisms that make that skill possible. The work we present here is the first stage of a project designed to resolve the major obstacles that have retarded the study of imitation and related motor behaviors.

Among the difficulties that have hindered study of imitation is the absence of appropriate, controlled but flexible test materials, and the challenge of properly quantifying the fidelity with which behaviors are being imitated. As imitated behaviors increase in complexity

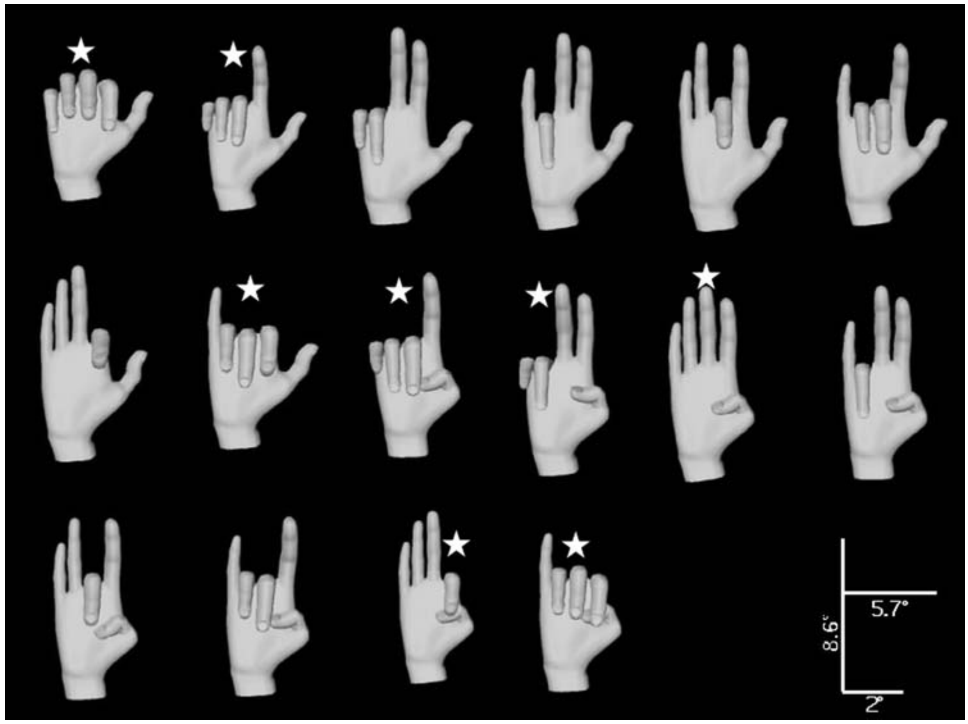

Fig. 1 The 16 component gestures used in generating each 6-item sequence of gestures. Gestures that match or closely approximate letters in American Sign Language are starred

(i.e., imitating a series of component actions), the obstacles hindering the study of imitation increase as well. In fact, to circumvent these obstacles, one influential study of imitation went to the lengths of limiting its test materials to displays in which just an index finger or a middle finger was flexed; that same study adopted an equally restricted method for quantifying imitation success, using only a binary, pass–fail scale (Iacoboni et al. 1999). And even when more complex behaviors have been examined, such as in studies of apraxics (Kimura 1993), assessment of imitation fidelity has incorporated substantial subjectivity, and therefore might fail to capture important aspects of performance.

Formally, the problem of quantifying the match between a model and an imitation of that model can be described as assessing the similarity of one n-dimensional trajectory to another. Drawing on examples from many domains other than imitation, Vlachos et al. (2002) catalogued the significant challenges in comparing two trajectories in any domain. Many of these challenges are relevant to the study of imitation, particularly in the common case where an imitation deviates from its stimulus model in spatial and/or temporal parameter. Such disparities foreclose the use of simple Euclidean metrics for comparing imitation and model, because any of these disparities would be represented in time series of unequal length. In addition, a simple Euclidean metric, even if it could be applied, would assign uniform weights to all the values in the n-dimensional time series representing the model and its imitation. Without significant embellishment, this uniform weighting would elide numerous theoretically important characteristics, including sequencing and serial-order effects (e.g., Lashley 1951; Agam et al. 2005; Agam and Sekuler 2007).

In the initial application of our analytic method we examined imitation of sequences of gestures, each drawn from a pool of 16 different patterns of digit flexions and extensions. These flexion/extension patterns, which are shown in Fig. 1, were selected as being representative of a range of other motor behaviors, and because they lent themselves to controlled variation and recombination in many different, novel sequences. These qualities have been important in devising test materials for other research domains, including memory (Ebbinghaus 1885/1913). Finally, we were attracted to these stimuli because of their kinship to gestures in American Sign Language (ASL), which we will study as part of our overall project (Fig. 1).

We present some experimental data that demonstrate our approach. Results show that our method can successfully generate a useful multivariate description of various types of errors made during imitation of gesture sequences, and can characterize practice-based improvements in imitation.

Method

Subjects

Eight paid subjects were recruited from the Brandeis University community. The subjects, whose ages ranged from 19 to 29 years old ($M = 21.25$, $SD = 3.05$), reported no prior experience with ASL; this exclusion criterion reflected the fact that some of our stimuli were similar to letters in ASL's finger spelling alphabet. All subjects provided written informed consent for the experiment in accordance with the principles of the Declaration of Helsinki. The experimental protocol had been approved by Brandeis University's Committee for the Protection of Human Subjects. All subjects had normal or corrected to normal vision and were strongly right-handed as determined by the Edinburgh Handedness Inventory (Oldfield 1971).

Materials

Using the method described below, model sequences were generated from the set of 16 gestures shown in Fig. 1. The Vizard VR Toolkit (WorldViz, Santa Barbara, CA) displayed the stimulus sequences on a 21″ CRT monitor with a refresh rate of 85 Hz. The stimulus was a right hand displayed on the screen with the palm facing the subject. At its longest, from the wrist to the tip of the middle finger, the stimulus model hand was 8.6° visual angle. The width of the model's wrist and palm, with all digits extended, were 2° and 5.7° visual angle, respectively.

Allowing all five digits of one hand to be either flexed or extended produces a set of $2^5 = 32$ hand gestures. Because we suspected that all 32 gestures would not be equally easy for subjects to reproduce (Schieber and Santello 2004), we carried out a preliminary study to identify a set of gestures that would be biomechanically possible for every subject to reproduce, and of approximately equal difficulty to reproduce. Six subjects viewed each of the 32 gestures four times in a random fashion. None of these subjects would serve in the actual experiment. After each observation, subjects attempted to reproduce the gesture and then used a scale from "1" (very easy) to "5" (extremely difficult) to rate how difficult the gesture was to reproduce. This rating constituted a *self-report* of difficulty. In addition, the data from each trial were analyzed to determine if the subject formed the correct hand position (*behavioral performance*).

Flexion values were measured for each digit ranging from 0 (completely extended) to 1 (completely flexed). A digit was categorized as flexed if the flexion value exceeded the 0.5 threshold, and extended if the flexion value fell below the 0.5 threshold. Each digit reproduction was compared to the corresponding model digit and judged to be correct (0) or incorrect (1). Thus, if all the digits in the reproduction were correct, subjects received a score of 0, whereas a reproduction with each digit incorrect would receive a score of 5. Data were averaged for each of the 32 gestures, producing an average score for each gesture for both behavioral performance and self-report. To ensure that the gestures used in our experimental protocol were all biomechanically possible for subjects to perform and of approximately equal difficulty we selected gestures with an average self-report score of less than 2, and an average behavioral performance of less than 0.5. This resulted in a total of 16 gestures for use in the demonstration experiment.

Apparatus

Subjects performed their imitations while wearing a right-handed, one-size model of the 5DT DataGlove 5 Ultra (Fifth Dimension Technologies) along with a hand sensor and a lower arm sensor from the Patriot™ motion tracking system (Polhemus). Rather than measuring the Cartesian coordinates of digit endpoints, for each digit, the data glove measures the flexion/extension of the intermediate and proximal phalanges of each digit. To determine the maximum flexion and extension of each digit, each subject must perform a short series of calibration routines where they are required to flex and extend each digit. The system then normalizes the data received from the glove in such a way that the maximum flexion of each digit is set to 1, and the maximum extension is set to 0. Subjects were instructed to make their movements naturally and not to overextend their hands or to flex their digits too tightly (i.e., they were instructed to make a loose fist as opposed to forcing all five digits into their palm). Hand and lower arm sensors measured the position of the hand and arm in the x, y and z dimensions as well as measuring yaw, pitch and roll. However, as the position of the hand and arm in the demonstration experiment were kept constant, our current analysis is limited to data collected from the data glove.

Stimulus construction and description

Sequences were generated by a Matlab program whose input was an ordered set of n gestures, and whose output was a seamless sequence in which each individual gesture blended smoothly into the next. The time for which each static gesture was held, as well as the transition times between successive static gestures were based on pilot observations with subjects who were excluded from the demonstration experiment itself. With these pre-defined hold and transition times for gestures, our sequence-generating software used tweening[1] to interpolate 42 frames

[1] Tweening is an animation technique that interpolates the differences between two existing key frames in an animation timeline. Tweening can operate on differences between the pre-existing frames, in attributes such as scale, opacity, location, color and shape.

between gestures in the sequence, blending successive static gestures with seamless transitions, as suggested in Fig. 2b. As a result, when the entire sequence of frames is displayed, an observer experiences smooth movement from the *n*th gesture in the sequence, through the tweened frames, to the $n + 1$th gesture in the sequence. In the resulting sequences, transitions from one gesture to another were smooth, biomechanically possible and natural seeming.

We used three types of stimuli: static gestures, two-gesture sequences, and six-gesture sequences. The first two types were used as practice stimuli to familiarize subjects with the gestures; the third constituted the experimental stimulus materials. The 16 different 6-item stimulus models in the experiment proper were divided equally between two conditions, which were designed to manipulate the cognitive demands required for each sequence by varying the number of digits whose changed state had to be noted for each change in gesture. For *one-transition* gesture sequences, sequences were constructed so that only one digit changed (i.e., was newly extended or flexed) between successive gestures in each sequence. In sequences of the *two-transition* condition exactly two digits were made to change between consecutive gestures. In addition, each successive pair of gestures appeared just once in the eight different stimulus models of each transition condition.

As mentioned above, the two transition conditions were used to manipulate the complexity needed to imitate the sequences. With stimuli from the one-transition condition, a total of five digits could change in flexion/extension from the first gesture to the sixth; with stimuli from the two-transition condition, a total of ten digits could change from the first gesture to the sixth. As a subject would have to remember and perform twice as many items (digit states) in the two-transition condition than in the one-transition condition, we hypothesized that imitating two-transition models would be more difficult for subjects to imitate than the one-transition models, and that, consequently, subjects would make more errors in imitating two-transition sequences. We included these two distinct types of sequences to verify that our algorithm could recover the performance difference expected from the two conditions.

Every gesture sequence, regardless of condition, began and ended with an open hand. These "bookend" open hand gestures were added to embed the six actual experimental gestures in comparable contexts. Without the bookend gestures, the first and the last of the six experimental gestures would lack a pre- or post-transition, respectively, so that the timing of their imitation could not be compared to the other four gestures. The subjects' reproduction of the bookend gestures was excluded from all analyses.

Procedure

Subjects viewed the stimuli while seated at a table, right elbow on a wrist rest and forearm and digits extended straight up. The palm faced the subject, just as the stimulus was displayed with the stimulus palm facing the subject. As a result, there was no need for the subject to perform a coordinate transform. For each 6-item sequence, stimulus presentation took 11.5 s. After the presentation completed,

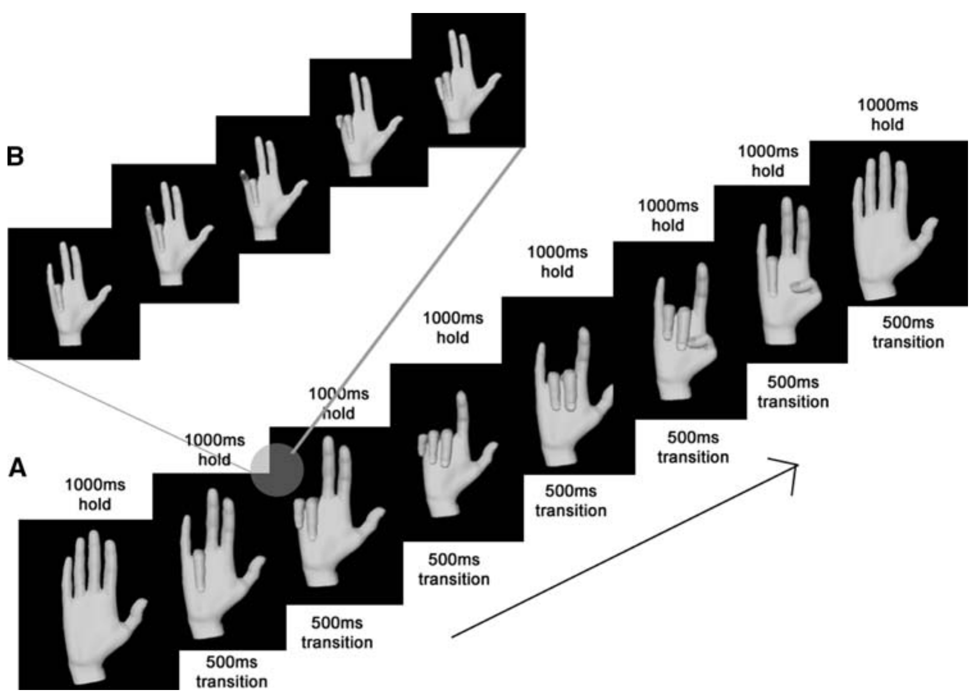

Fig. 2 **a** An example of one 6-item stimulus model. Note that the model begins and ends with all digits fully extended, and that the change from one gesture to the next involves a flexion or extension of exactly one digit per gesture. **b** The inset shows five sample digit postures that were transitional movements between the second and third gestures in the sequence. Forty-two transitional digit postures were shown between each pair of successive gestures in each sequence, resulting in a very smooth movement throughout the entire sequence

the screen cleared, and subjects waited for a tone (2 s after the stimulus had completed) before beginning their imitation. Subjects were instructed to accurately imitate as many gestures in the sequence as possible, and try to attempt to reproduce them in the correct order. Subjects were allowed 14 s to complete their imitation of each 6-item sequence. No subject reported needing more time to complete their imitation. Though they could see their own hand while performing the imitation, no other explicit feedback about imitation accuracy was provided. After the response time elapsed, subjects were instructed to press a key with their left hand to start the next trial.

Calibration phase

At the start of the session, we calibrated the data glove by presenting a subject with images of four hand postures. These postures involved various configurations with the digits being flexed or extended, for example an open hand and clenched fist. During the 10-s period in which the images were visible, a subject reproduced them in succession, as many times as possible. Subjects were monitored to ensure that they reproduced each calibration hand posture at least once.

Familiarization phase: I

To begin the process of familiarizing subjects with each component gesture they would later see in the multi-item gesture test sequences, subjects viewed and imitated each of the 16 static, component gestures that would appear in those sequences. Subjects started each trial with an open hand, with their palm facing them. Each gesture, shown in Fig. 1, was displayed for 1 s. Two seconds after the gesture disappeared from the display screen, the subject heard a tone indicating that they were to reproduce the stimulus gesture from memory. Subjects were instructed to initiate their movement as quickly as possible after the "go" signal and to hold the gesture for 2 s (when a written instruction indicated that they should return to the starting position). All 16 gestures were presented twice, in block randomized fashion.

Familiarization phase: II

Subjects next viewed and imitated eight different two-gesture sequences. Each sequence began and ended with an open hand, and each gesture was held for 1 s, with a transitional time of 500 ms. Thus the total stimulus presentation time for these two-gesture sequences was 5.5 s. As in the first familiarization phase, 2 s after the gesture disappeared from the display screen, the subjects heard a tone indicating they were to reproduce the sequence from memory. They had a window of 7 s in which to reproduce the sequence. Overall, these eight familiarization sequences incorporated all 16 component gestures. As each sequence was presented just once in this phase, over both familiarization phases subjects saw and imitated each of the sixteen gestures three times.

Experimental phase

Finally, in the experimental phase of the procedure, each subject viewed and imitated eight different 6-item gesture sequences. Each sequence began and ended with an open hand, and included tweened frames between successive gestures. Each gesture was held for 1 s, with a 500 ms transitional time between gestures. The eight model sequences were shown and imitated in massed fashion; that is, each model was viewed and imitated 10 times before the subject saw the next model sequence. Additionally, and unbeknown to subjects, the experimental phase was divided into two equal parts, with each part containing four models from one of the transition conditions (i.e., one-transition models or two-transition models). Thus, half the subjects imitated four of the two-transition models followed by four of the one-transition models, while the remaining subjects imitated four of the one-transition models followed by four of the two-transition models. The choice of models from each transition condition was fully counterbalanced across subjects.

Segmentation

To compare subjects' imitations of gesture sequences to the corresponding model sequences, we developed a multi-stage algorithm that determined the differences between the model and the subject's reproduction. As both the model and the imitation sequence contain a specific number of component gestures, to compare the imitation sequences to the model sequences, the algorithm first needs to segment both the imitation and model and then compare the corresponding individual components. Toward this end, the algorithm resamples the data to achieve a constant sampling rate of 50 Hz. Each resulting sample $x(t)$ is represented by a 5-component vector containing the flexion values, ranging from 0 to 1, of all digits at time step (or sample number) t. Then it computes the combined velocity $v(t)$ of digit motion at a given time step t by computing the root mean square flexion difference in the five digits occurring within a narrow time window centered at time t. Instead of simply measuring the changes between consecutive samples, this window was introduced to reduce the noise in velocity measurement. The ideal size of the time window depends on the accuracy and frequency of flexion measurement. In the present study, the window was set to 140 ms, or 7 samples—from $(t - 3)$ to $(t + 3)$:

$$v(t) = \frac{1}{f} \times \sqrt{(x(t+3) + x(t+2) + x(t+1) - x(t-1) - x(t-2) - x(t-3))^2}$$

where f is the number of flexion values collected per sample, so $f = 5$ in the present study. In its next stage, the algorithm examines the velocity information throughout a given imitation and categorizes successive intervals in the imitation as one of two possible types of components: gestures (static or held) and transitional movements (movements from one static gesture to another). A component gesture is defined as static when the velocity drops below 10% of its peak velocity and remains below that level for at least 100 ms; the duration threshold of 100 ms was empirically determined to yield the most plausible results for the present type of motion sequences. A transitional movement is defined by an interval between two consecutive component gestures in which the combined velocity does not fall below 10% of the peak velocity for a duration of 100 ms or longer.

Comparison to model stimulus

Once the transitional movements and component gesture time epochs have been defined, our algorithm examines the flexion data for each digit in each time epoch, weighting each digit equally. The flexion threshold is defined as a value of 0.5, with any digit whose flexion value exceeded that threshold considered to be flexed, and any digit with a flexion value below the threshold considered to be extended. Each gesture is demarcated by the digits that were extended, starting with the thumb (1) and ending with the little finger (5). Thus, a completely open hand—with all digits extended—would be represented as **12345**, and a gesture with only the middle and ring fingers extended would be represented as **34**. Figure 3 shows the results of the segmentation process for one trial, as well as the pictoral gesture representation of the raw data below the graph.

Once the component static gestures have been identified by our algorithm, the subject's imitation is compared to the original model, in both accuracy and timing. The bookend gestures (i.e., open hand gestures) are first removed from both the model and the reproduction. Each gesture in the imitation is compared to each gesture in the model and the algorithm reorders the imitation to minimize the sum of incorrectly reproduced digits across all gestures in the sequence. The total number of errors in an imitation are calculated from the sum of the values of different error categories.

Categories of errors

When imitating sequences of hand gestures, there are three main categories of errors that subjects could commit, namely gesture-level, sequence-level, and unmatched errors.

Gesture-level errors

Gesture-level errors are defined by one or more flexion differences between a reproduced gesture and the model gesture that it was meant to reproduce. There is currently a divide in the scientific community as to the existence of digit somatotopy in the hand's representation in primary motor cortex (M1), though many studies suggest a considerable overlap of digit representations in M1 (see Sanes and Schieber 2001 for a summary, and Hlustík et al. 2001; Rao et al. 1995; Sanes et al. 1995 for studies on the overlap). Despite the controversy, human hands have well-defined biomechanical constraints, such that when an intended digit is in transition between flexion and extension (or vice versa), the adjacent digits move more than the non-adjacent digits do (Fish and Soechting 1992; Häger-Ross and Schieber 2000). As a result, our algorithm was designed to differentiate between (1) an interchange of adjacent digits, and (2) an interchange of non-adjacent digits, with biomechanical coupling making the former kind of interchange more likely than the latter (Loehr and Palmer 2007; Schieber and Santello 2004). An adjacent gesture-level error is the flexion swapping between two adjacent digits. For example, if the model **124** were reproduced as **125**, then the subject simply extended the little finger instead of the ring finger. A non-adjacent gesture-level error involves an interchange of two non-adjacent digits, such as the index and ring fingers in the reproduction of **12** for **14**.

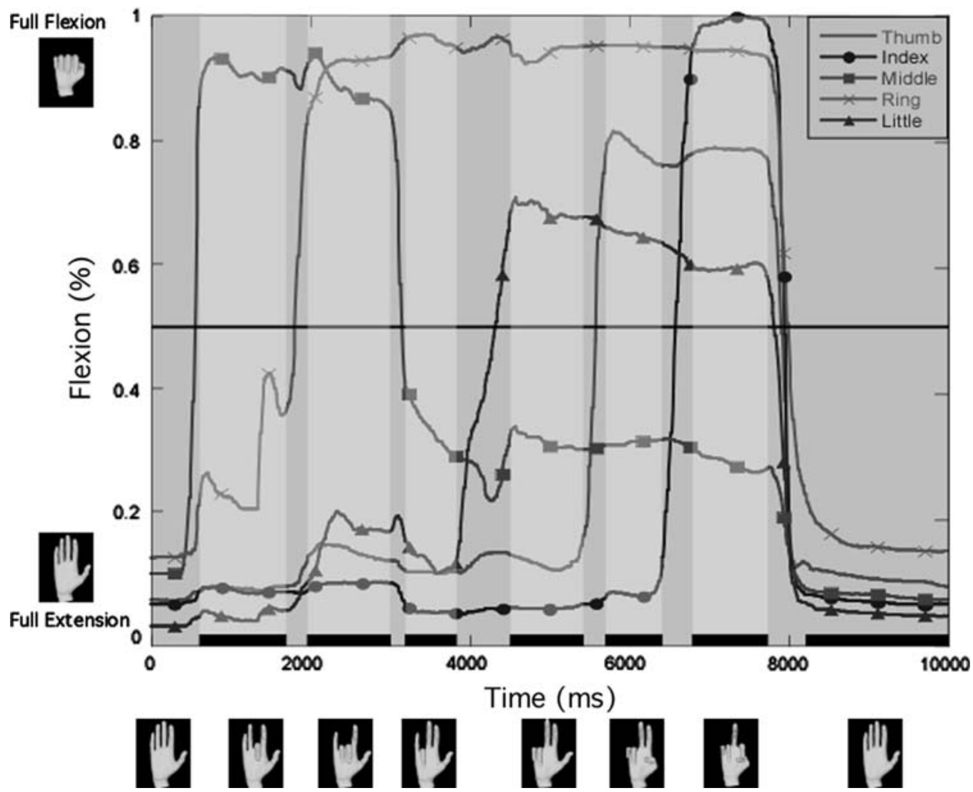

Fig. 3 An example of the segmentation stage of the algorithm. *Light gray vertical bands* indicate individual static gestures as determined by the segmentation process. *Dark gray vertical bands* indicate transitory movements. The *vertical line* at $y = 0.5$ is the flexion threshold. The *black horizontal lines* along the *x*-axis, below each *light gray* band, indicate the duration of each static gesture

In both of the above cases, the number of extended digits in the model is correctly reproduced. However, there are other gesture-level errors that do involve a mismatch in the number of flexed digits between the model and the imitation. Consider, for example, an error that we call an individual flexion error: the gesture **234** is reproduced as **24**, or conversely, the gesture **24** is reproduced as **234.** In the first of these individual flexion errors, one digit extension has been omitted (an Omission Error); in the second of these individual flexion errors, an extra digit extension has been inserted (an Insertion Error). Figure 4 shows an example of a gesture-level error where, for the sixth gesture, the subject produced **3** instead of **234**.

Sequence-level errors

At the sequence level, similar types of errors can occur. As gestures in our task are performed sequentially, there are obvious parallels to work on serial order with verbal materials: performance in both domains shows evidence of failures in working memory (Conrad 1960; Lee and Estes 1977; Lewandowsky and Murdock 1989). A common sequence-level error is the mistake of swapping two consecutive gestures in a sequence. For example, if the sequence **123–1234–134–13** were reproduced as **123–134–1234–13**, then such an error is represented by the second

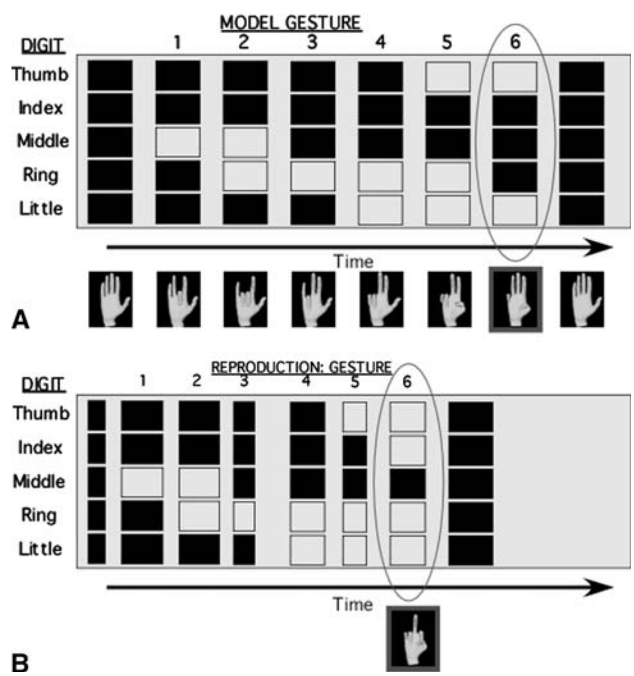

Fig. 4 **a** The model stimulus to be imitated. **b** A subject's imitation of the model stimulus as seen in **a**. As shown, the subject correctly imitated the first five gestures (the first and last gestures, the open hand, do not count). But, with the sixth gesture (highlighted in both the model and subsequent imitation), instead of extending the ring finger, the subject flexed the index finger. In addition, the timing is slightly different, as **a** and **b** are time-scaled comparisons

and third reproduced gestures, whose order has been inverted. In some cases however, two non-consecutive gestures may be swapped, such as gestures number two and four when the model **235–245–24–23** is reproduced as **235–23–24–245**. It is also possible that the sequence is permutated in such a way that there is no pair of gestures whose positions were exactly swapped. For instance, let us assume the model **345–34–45–134** and its reproduction **134–34–345–45**. Here, we have no pairwise swapping, but three of the four gestures (numbers one, three, and four in the model) occupy incorrect positions within the reproduced sequence. These are counted as three sequence-level errors. We can assign sequence-level errors to individual gestures in the model sequence, that is, we can determine whether a particular gesture was imitated correctly with regard to digit flexion and serial position. This allows us to study the accuracy of gesture imitation as a function of the gesture's serial position in the model sequence.

Unmatched errors

Finally, it is possible that the algorithm is unable to match an imitated gesture to one in the model. There are two possible causes to this, both involving a mismatch in the number of gestures reproduced versus the number of gestures in the model. First, the imitation may contain more gestures than the model. In this case, after all the model gestures have been matched with those from the imitation, the number of extraneous, unmatched gestures in the imitation is the number of unmatched errors in that imitation. For example, if the sequence **123–1234–134–13** were reproduced as **123–1234–134–1234–13**, the algorithm will find the second **1234** is an extraneous gesture, and there will be one unmatched error in the imitation. Second, a subject might reproduce fewer gestures than in the model. In this case, all the gestures that are imitated will be best matched to the model; any leftover model gestures will comprise the unmatched errors. For example, if the sequence **123–1234–134–13** were reproduced as **123–13**, the algorithm will find that **1234** and **134** were not reproduced, which will result in two unmatched errors.

In identifying the types and positions of errors in some reproductions, we run into a problem of ambiguity: In many cases there are two or more error patterns that could have caused the observed discrepancies between model and reproduction. To illustrate the ambiguity within an individual gesture, consider a model gesture **234** and its reproduction **345**. Clearly there is an error, but that error could have arisen from a non-adjacent flexion swapping of the index and little fingers, or two individual flexion errors for the same fingers. If we include errors at the sequence level, the situation becomes more complex, as illustrated by the following example: Model **1234–125–12** is reproduced as **123–125–1234**. One possible explanation of the underlying errors is that one individual flexion error occurred in gesture 1, and two of the same type occurred in gesture 3, while the order of gestures was correctly reproduced. Alternatively, the subject may have swapped gestures 1 and 3 during imitation, and introduced an individual flexion error when reproducing **12** so that it became **123**. Other error patterns as well could have caused the observed reproduction.

To reduce ambiguity and determine the most plausible underlying error pattern, the algorithm first attempts to put the imitation in a sequence that best reflects the original model; that is, it reorders the gestures in the imitation in order to minimize the sum of the incorrectly reproduced digit flexions across all gestures in the sequence. Next, the algorithm looks at the differences between flexions in each imitated gesture and flexions in the model gesture. The observed differences in the flexion of each gesture comprise a minimum number of elementary errors (swapping, inserting, and omitting flexions). Finally, any extraneous gestures in the imitation that are not matched to the model and any gestures that are in the model but not matched to the imitation comprise the unmatched errors.

Statistical analysis

When subjects failed to start and finish a sequence with an open hand, that trial was excluded from our data analysis. A failure to start the sequence with an open hand suggests that subjects initiated their response before the "go" cue. A failure to end a sequence with an open hand suggests that subjects either forgot to return to the open hand, or they ran out of time to complete their imitation. Of all experimental trials, 5.625% were excluded for this failure.

Dependent measures included spatial errors and temporal errors, including number of gestures reproduced, total errors, gesture level errors, sequence level errors, unmatched errors, serial position errors, pre-movement latency, movement time and average transition time between segments. All statistical analysis was performed with SPSS. Each dependent variable was subjected to a repeated-measures ANOVA, with condition (one-transition or two-transition) and repetition (1–10) being within-subject variables. For serial position errors an additional within-subjects factor of serial position was incorporated into the analysis. Where sphericity assumptions were violated, Hunyh–Feldt corrections were applied. A significance threshold of 0.05 was used throughout. For conciseness only significant findings are reported here.

Results

Number of gestures produced

Subjects were told to reproduce as many gestures as they could remember, but that number of reproduced gestures often fell short of the number (six) that comprised the model. The average number of gestures that subjects produced in their imitations did not differ significantly between the conditions of transition ($F_{1,7} = 1.489$, $p = 0.262$). Repetition of a model significantly influenced the number of gestures reproduced, with that number increasing systematically from the first to the tenth presentation ($F_{6.777,47.437} = 9.909$, $p < 0.001$). This effect can be seen in Fig. 5a. There is also a significant interaction between transition condition and repetition ($F_{9,63} = 5.910$, $p < 0.001$), with a more consistent number of gestures reproduced in the one-transition condition. These effects need to be considered in the computation of error scores, as the more gestures a subject reproduces, the greater is the opportunity for gesture- or sequence-level errors. Therefore, data were normalized according to the number of gestures that were produced.

Spatial errors in imitation

The (normalized) total number of errors decreased with repetitions, demonstrating an improvement in performance with practice ($F_{4.861,29.165} = 4.853$, $p < 0.001$). Further, we see more errors in early repetitions in the two-transition condition compared to the one-transition condition, with a significant interaction between transition condition and repetition ($F_{5.642,33.849} = 3.945$, $p < 0.01$). Because our algorithm breaks the errors down into various categories (gesture-level, sequence-level, and unmatched gestures), we can look at these same effects for each error category. Figure 5b, c shows the normalized errors for the total errors and across each error category for the one- and two-transition conditions, respectively.

For gesture-level errors, subjects make less errors as repetition number increases ($F_{6.635,39.811} = 6.093$, $p < 0.001$).

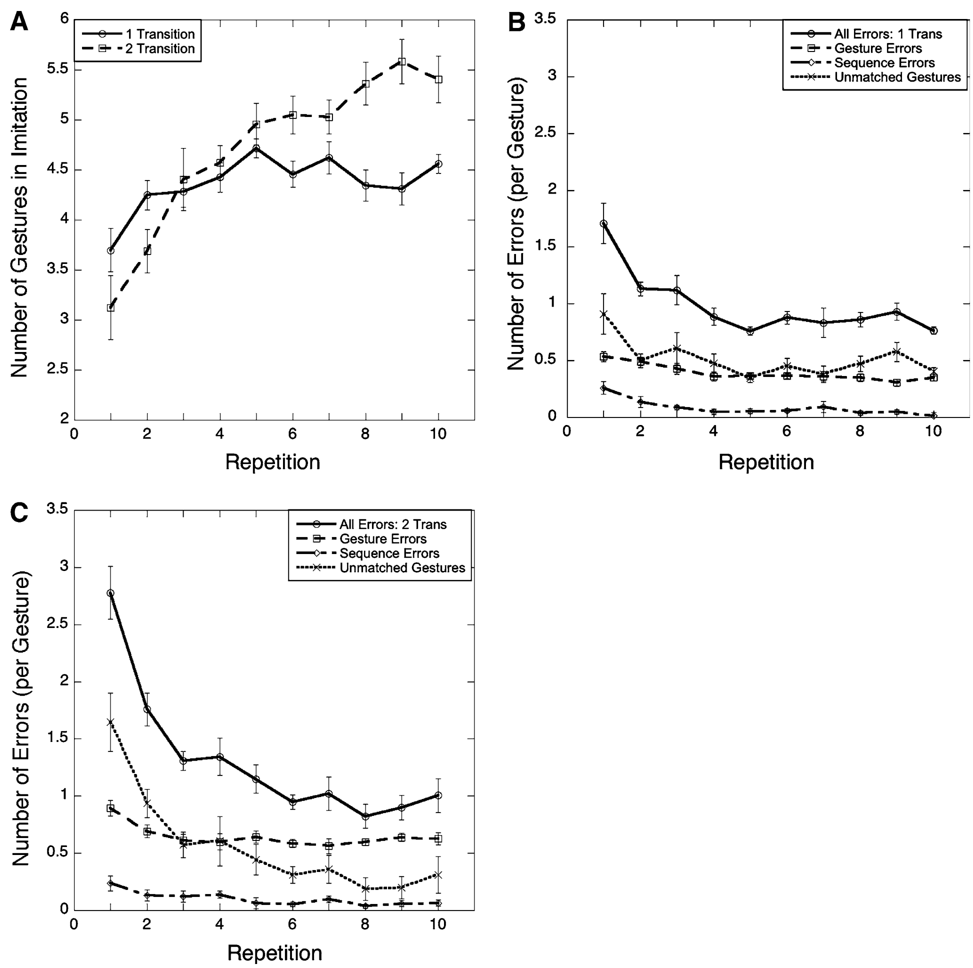

Fig. 5 a Mean number of gestures made in imitation for each repetition of a model sequence. **b** Mean number of errors, by error type, over repetitions of a one-transition model sequence; results are normalized by the number of gestures made in imitation. **c** Mean number of errors, by error type, over repetitions of a two-transition model sequence; results are normalized by the number of gestures made in imitation. In all panels, *error bars* represent within-subject standard errors of the mean

In addition, subjects make approximately twice as many gesture level errors in the two-transition condition compared to the one-transition condition ($F_{1,6} = 41.792$, $p = 0.001$). This is expected, given that twice as many digits change in the two-transition condition than in the one-transition condition. For sequence-level errors, subjects once again produce fewer errors with increased repetition ($F_{3.973,23.836} = 4.278$, $p = 0.010$). The same pattern is observed for the unmatched gestures, with improvements occurring with repetition ($F_{5.634,33.806} = 6.559$, $p < 0.001$). In addition, in the two-transition condition (compared to the one-transition condition), subjects make more unmatched errors during early repetitions, with this difference disappearing in later repetitions; this is reflected in a significant interaction between repetition and transition condition ($F_{6.385,38.309} = 3.554$, $p < 0.01$).

Serial order

How accurately do subjects imitate the serial order of the model? That is, is a specific gesture in the imitation correctly matched to one in the model in both order in the sequence and in digit flexion? To calculate serial position we compared each item in the model sequence to each corresponding item in the reproduction, a correct match was assigned a value of 1 and an incorrect match was assigned a value of 0. We find that errors decrease with repetition ($F_{5.556,38.889} = 4.828$, $p = 0.001$) and increase with serial position ($F_{3.125,21.874} = 19.284$, $p < 0.001$), in other words, during the first gestures subjects make fewer mistakes. This primacy effect is significantly larger for the one-transition condition compared to the two-transition condition (Fig. 6a, b, respectively), shown by a significant interaction between transition condition and serial position ($F_{4.291,30.035} = 4.493$, $p < 0.01$).

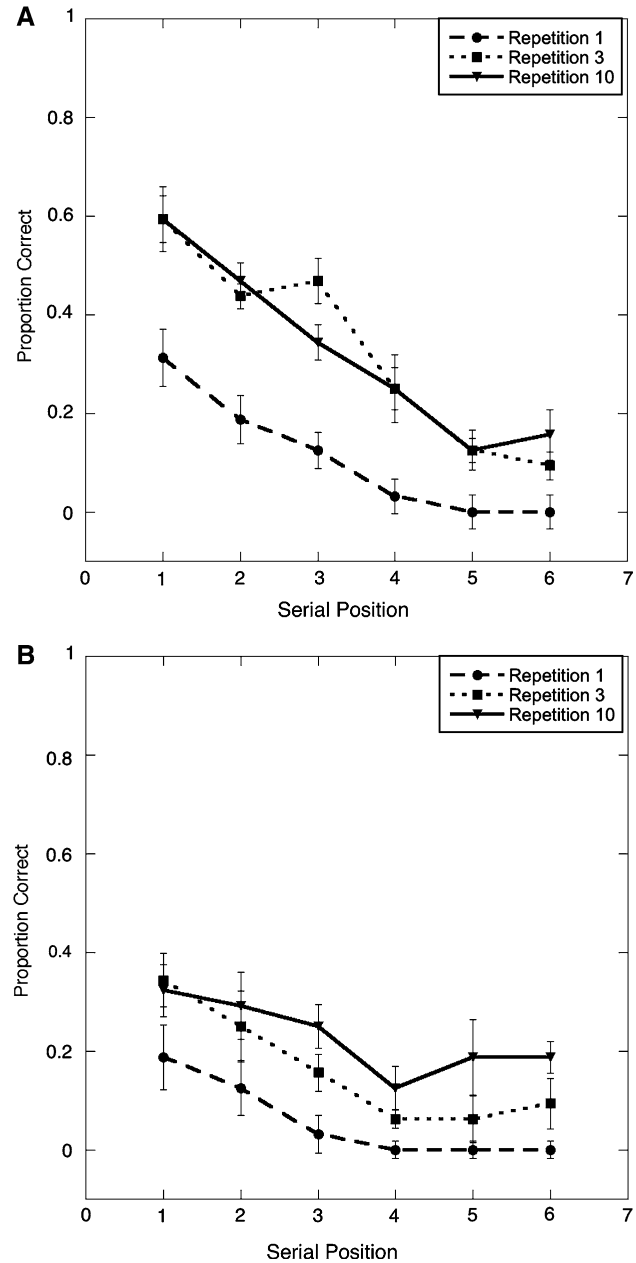

Fig. 6 Accuracy as a function of item position in the model sequence for multiple repetitions for the one-transition (**a**) and two-transition (**b**) models

Temporal analysis of imitation

Pre-movement latency is the amount of time a subject holds the initial open hand before initiating the transition to the first gesture in the imitation, after the tone sounds. Subjects show a decrease in pre-movement latency from repetition one (M = 1,200.938, SD = 457.304) through to repetition ten (M = 798.438, SD = 268.301), indicating that they take significantly less time to prepare their response as they become more familiar with the stimulus ($F_{9,54} = 3.243$, $p < 0.01$).

The movement time is defined as the total amount of time the subject takes to complete the imitation, excluding the time both open hands are held. As this is highly dependent on the number of gestures produced, movement time is expressed as a function of the number of gestures reproduced. We observed a decrease in movement time as repetition number increases from repetition 1 (M = 2,149.734, SD = 460.551) to repetition 10 (M = 1,645.175, SD = 334.742); this effect is statistically significant ($F_{6.164,36.981} = 4.089$, $p < 0.01$). In addition, subjects take longer to perform the two-transition stimuli (M = 1,987.395, SD = 376.572) than the one-transition stimuli (M = 1,555.560, SD = 323.188); this main effect of transition condition is also significant ($F_{1,6} = 6.533$, $p < 0.05$).

The mean transition time is the average of all the transition times between static gestures. Transition times are much slower for two-transition models (M = 806.342, SD = 96.107) than for one-transition models (M = 509.118, SD = 26.551), a difference that is statistically significant ($F_{1,7} = 12.088$, p = 0.01).

Discussion

We have presented a novel methodology for assessing the fidelity of human imitation of complex movements. This methodology builds on a recent scheme introduced by Agam and colleagues who measured the fidelity of imitation using an automated segmentation of behavior sequences into constituent parts (Agam et al. 2005, 2007). Their scheme incorporated the same spatio-temporal discontinuities that human observers use when segmenting continuous behaviors into constituent subcomponent behaviors (Zacks and Tversky 2001; Zacks et al. 2001). Though the analytic technique used by Agam et al. (2005) succeeded in opening important insights into the neural mechanisms that support imitation, that technique's scope is limited to a narrow range of stimuli and responses, namely linked, linear, two-dimensional motion sequences. We address that limitation here by presenting a novel, far more flexible method that generates a multivariate assessment of imitation quality for complex, realistic movements.

Our approach compares an imitation of any multidimensional sequence to the original model sequence on a multivariate level. This enables the study of sequence-based motor learning to move beyond simple comparisons, such as subjective analysis or using a pass–fail measure, as we are now able to examine more intricate properties of sequenced behavior. Using the velocity components of an imitation, the algorithm segments the movement of the model and the imitation into component parts and then makes a spatio-temporal comparison of individual components of the model sequence to the imitated sequence.

Here, we have presented data that illustrate the effectiveness of our algorithm. We asked subjects to imitate sequences of hand movements, which were repeated ten times. We showed that our algorithm could successfully isolate various types of spatial errors and quantify them. We demonstrated that the total number of errors decreases with repetition, and that different error types can account for the total number of errors (with omissions and insertions accounting for most of our error types). As we are able to break an imitation down into component parts, our algorithm also allows us to compare our results to those from studies in other domains, such as short-term visuospatial memory. In the one-transition condition, we found both a primacy effect and a trend towards a 1-item recency effect, similar to findings from other studies of short term memory, using either verbal materials (Lee and Estes 1977; Lewandowsky and Murdock 1989) or non-gestural motor imitation (Agam et al. 2005, 2007). Interestingly, even after ten repetitions of each sequence in massed fashion, there is little change in the shape of the serial order curves seen in Fig. 6, a phenomenon that rules out several theoretical accounts of practice-based imitation learning (Agam et al. 2007).

In addition, we illustrated that our algorithm could identify chronometric properties of imitation, with longer pre-motor latencies observed during early imitations in subjects compared to later imitations. This presumably reflects a learning effect in that the cognitive effort needed by the subject decreases with repetition, as the subject learns the sequence. To further advance this claim, we also found an effect of repetition on the total time to complete the imitation (normalized by the number of gestures produced), with subjects performing the imitation faster with repetition.

Our algorithm also recovers the differences in performance that arise from differences in the complexity of the two different transition conditions. Subjects perform the one-transition condition faster than the two-transition condition. This effect is further supported by the finding that the transitions between static gestures are slower in the imitation of the two-transition sequences when compared to those of the one-transition sequences, and that there are about twice as many gesture-level errors in the two-transition condition than in the one-transition condition. In addition, the difference in sequence complexity is also reflected in many of our spatial error measures (including number of gestures reproduced, total number of errors, total number of unmatched errors, and serial position errors), with subjects in the two-transition conditions showing more errors on early repetitions, with the two conditions converging on later repetitions when subjects have had more practice at the sequences.

In this paper we focused on the analysis of imitation of gesture sequences that were observed, stored and retrieved from memory. However, it is important to recognize that the analytic tools described here should be useful in other settings as well. In the work reported here we segmented trajectories on the basis of finger flexion velocity, but this same segmentation can easily be extended to include more factors. For example, a study involving imitation of naturalistic arm movements could use the methodology presented here to segment the arm's trajectory according to the velocity profiles of not just the fingers, but the upper arm and forearm as well. Thus, the algorithm could be easily extended to include any or all components reported by a motion capture device. As a result, our methodology is not limited to simply finger flexions, or even to descriptions based on Cartesian coordinates of upper and lower arms, in the case of a reaching task; it can characterize relative or absolute angles, as well as the roll, pitch, and yaw of the various components. Using the resultant segmentation points, our algorithm could quantify and categorize types of errors in performance, and capture the chronometric properties of performance, assigning appropriate weights to various components.

Its extensibility allows our basic approach to be applied to the analysis of complex movements, producing useful information, including chronometrical information, which is difficult to achieve with ordinary video-based analysis and description of gestures. To take one example, data from patients with a motor system deficit might resist analysis because of idiosyncratic and seemingly random movements and movement timing that may intrude during an attempted imitation. To deal with such anomalies, post-segmentation parameters of the analysis could be fine-tuned in various ways, e.g., adjustment could be made to the duration threshold that we used here to define a gesture component (100 ms). Additionally, the same method could search through the motion capture data for an imitation that correctly matches just a portion of the model trajectory. The latter is especially important for patients who, in the process of completing their imitation, actually imitate the model with fair fidelity, despite incorporating extra components before or after the imitation itself.

The novel methodology presented here affords a way to answer many questions about the production of complex movements, and about the contribution of the various components of the human praxis system. The praxis system represents a fronto-parietal network of brain regions dedicated to tool use. Damage to this system results in apraxia, an inability to perform purposeful skilled movements. With one form of apraxia, ideomotor apraxia, patients can make spatial and temporal movement errors when asked to pantomime tool use, gesture to command, imitate movements or use real tools and objects (Heilman and Rothi 2003). With only modest changes, our analytic tools can be adapted to identify subtle deficits in pantomimed tool use performance (Halsband et al. 2001; Sunderland and Shinner 2007), as is often used in the diagnosis and evaluation of apraxic patients. In this context, pantomimed tool use refers to an instruction to demonstrate how one would use a tool or object, either in the absence of the object or without making contact with the object. For example, a subject may be given a command such as “Show me how you would use a hammer.” In a series of recent experiments, the movement errors made by ideomotor apraxic patients were assessed by scoring a videotape of subjects’ performance (Haaland et al. 2000; Halsband et al. 2001; Buxbaum et al. 2005; Rumiatai et al. 2005; Jax et al. 2006; Sunderland 2007). Although that approach may be adequate in identifying gross errors, it lacks the ability to precisely define and identify various types of errors and the chronometric properties of such errors. Our novel methodology has the potential to overcome such shortcomings.

Another potential use of the methodology we have described herein, is the possibility to examine spatial and temporal errors produced by individuals with unusual expertise or a suspected deficit in some movement-related domain. The former would include individuals whose gestural expertise arises from their fluency in ASL, particularly as about half the gestures used in our sequences very closely approximate letters from the ASL alphabet (for example, Whitehead et al. 1997; Jerde et al. 2003). Individuals with suspected deficits would include those with some form of autistic spectrum disorder (Iacoboni and Dapretto 2006; Hamilton et al. 2007; Vanvuchelen et al. 2007). Finally, we believe that our analytical methods could be incorporated into systems that deliver automated, robot-assisted rehabilitation to post-stroke patients (Li et al. 2006; Mataríc et al. 2007). A detailed multivariate analysis of movements made by such patients could greatly increase the feedback that robot therapists give to post-stroke patients, which may well facilitate the rehabilitation process.

Of course, as with any new methodology, our technique arrives with *caveats*. To compare a subject’s series of hand gestures to that of a model sequence, our algorithm examines the flexion and extension of each digit of one hand, and characterizes the imitation’s various chronometric properties. In the experiment presented here, a pre-experiment calibration routine determined the maximum flexion and extension possible for each subject, and then normalized each subject’s data so that the maximum flexion was set to one, and the maximum extension was set to zero. We then employed a threshold of 0.5, categorizing a digit with a value of greater than 0.5 as being flexed, and a digit with less than 0.5 as being extended. Although this binary approach is suitable for our current experimental paradigm (where digits in the model were either fully flexed or extended), this approach would be less suited when model digits could assume more than just two states. In the near future, we plan to extend our basic method to encompass these more subtle cases, with additional states of flexion, and taking note of not only digit flexion, but also digit adduction.

It is also important to note that in addition to the flexion and extension of the digits, our existing sensors and methodology afforded the opportunity to examine the position and orientation of the hand and lower arm. However, for the demonstrations presented here, we held constant the position of the hand and arm, thereby restricting analysis to data collected from the data glove alone. We plan to extend our experimental paradigm to more complex movements by taking full account of hand and arm movements, as well as those of the digits. Finally, the results presented here came from a task in which subjects imitated the gestures of just a single hand. We plan to extend our approach to simultaneous movements of both hands, using tasks that require bimanual control and coordination.

In conclusion, here we have devised and tested a novel technique for assessing the fidelity of human imitation. We employed a behavioral paradigm to illustrate the efficacy of our algorithm at defining the components of the movement and for quantifying and categorizing errors in both the spatial and temporal domain. We feel this technique has broad implications for cognitive neuroscience and neuropsychology, allowing researchers and clinicians to ask previously unanswerable questions about the ability to imitate. We hope to extend this methodology to study different types of complex movements in subjects from special populations, such as experts and novices, as well as patients with deficits in the ability to program and perform certain complex movements.

References

Agam Y, Bullock D, Sekuler R (2005) Imitating unfamiliar sequences of connected linear motions. J Neurophysiol 94:2832–2843

Agam Y, Galperin H, Gold BJ, Sekuler R (2007) Learning to imitate novel motion sequences. J Vis 7:1–17

Agam Y, Sekuler R (2007) Interactions between working memory and visual perception: an ERP/EEG study. Neuroimage 36:933–942

Buxbaum L, Johnson-Frey S, Bartlett-Williams M (2005) Deficit internal models for planning hand-object interactions in apraxia. Neuropsychologia 43:917–929

Conrad R (1960) Serial order intrusions in immediate memory. Br J Psychol 51:45–48

Ebbinghaus H (1885/1913) Memory: a contribution to experimental psychology. Teachers College, Columbia University, New York

Fish J, Soechting JF (1992) Synergistic finger movements in a skilled motor task. Exp Brain Res 91:327–334

Haaland K, Harrington D, Knight R (2000) Neural representations of skilled movement. Brain 123:2306–2313

Häger-Ross C, Schieber MH (2000) Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies. J Neurosci 20:8542–8550

Halsband U, Schmitt J, Weyers M, Binkofski F, Grutzner G, Freund HJ (2001) Recognition and imitation of pantomimed motor acts after unilateral parietal and premotor lesions: a perspective on apraxia. Neuropsychologia 39:200–216

Hamilton AF, Brindley RM, Frith U (2007) Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? Neuropsychologia 45:1859–1868

Heilman K, Rothi L (2003) Apraxia. In: Heilman K, Valenstein E (eds) Clinical neuropsychology. Oxford University Press, New York

Hlustík P, Solodkin A, Gullapalli R, Noll D, Small S (2001) Somatotopy in human primary motor and somatosensory hand representations revisited. Cereb Cortex 11:312–321

Iacoboni M (2005) Neural mechanisms of imitation. Curr Opin Neurobiol 15:632–637

Iacoboni M, Dapretto M (2006) The mirror neuron system and the consequences of its dysfunction. Nat Rev Neurosci 7:942–951

Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999) Cortical mechanisms of human imitation. Science 286:2526–2528

Jax S, Buxbaum L, Moll A (2006) Deficits in movement planning and intrinsic coordinate control in ideomotor apraxia. J Cogn Neurosci 18:2063–2076

Jerde TE, Soechting JF, Flanders M (2003) Coarticulation in fluent fingerspelling. J Neurosci 23:2383–2393

Kimura D (1993) Neuromotor mechanisms in human communication (no. 20). Oxford University Press, New York

Lashley KS (1951) The problem of serial order in behavior. In: Jeffress LA (ed) Cerebral mechanisms in behavior. Wiley, New York, pp 112–136

Lee CL, Estes WK (1977) Order and position in primary memory for letter strings. J Verbal Learn Verbal Behav 16:395–418

Lewandowsky S, Murdock BB (1989) Memory for serial order. Psych Rev 96:25–57

Li ZM, Davis G, Gustafson NP, Goitz RJ (2006) A robot-assisted study of intrinsic muscle regulation on proximal interphalangeal joint stiffness by varying metacarpophalangeal joint position. J Orthop Res 24:407–415

Loehr JD, Palmer C (2007) Cognitive and biomechanical influences in pianists' finger tapping. Exp Brain Res 178:518–528

Mataríc MJ, Eriksson J, Feil-Seifer DJ, Winstein CJ (2007) Socially assistive robotics for post-stroke rehabilitation. J Neuroeng Rehabil 4:5

Oldfield R (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113

Oztop E, Kawato M, Arbib M (2006) Mirror neurons and imitation: a computationally guided review. Neural Netw 19:254–271

Rao S, Binder J, Hammeke T, Bandettini P, Bobholz J, Frost J, Myklebust B, Jacobson R, Hyde J (1995) Somatotopic mapping of the human primary motor cortex with functional magnetic resonance imaging. Neurology 45:919–924

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

Rumiati R, Weiss P, Tessari A, Assmus A, Zilles K, Herzog H, Fink GR (2005) Common and differential neural mechanisms supporting imitation of meaningful and meaningless actions. J Cogn Neurosci 17:1420–1431

Sanes J, Donoghue J, Thangaraj V, Edelman R, Warach S (1995) Shared neural substrates controlling hand movements in human motor cortex. Science 268:1775–1777

Sanes J, Schieber M (2001) Orderly somatotopy in primary motor cortex: does it exist? Neuroimage 13:968–974

Schieber M, Santello M (2004) Hand function: peripheral and central constraints on performance. J Appl Physiol 96:2293–2300

Sunderland A (2007) Impaired imitation of meaningless gestures in ideomotor apraxia: a conceptual problem not a disorder of action control?: a single case investigation. Neuropsychologia 45:1621–1631

Sunderland A, Shinner C (2007) Ideomotor apraxia and functional ability. Cortex, 43:359–367

Vanvuchelen M, Roeyers H, De Weerdt W (2007) Nature of motor imitation problems in school-aged males with autism: how congruent are the error types? Dev Med Child Neurol 49:6–12

Vlachos M, Kollios G, Gunopulos D (2002) Discovering similar multidimensional trajectories. In: Proceeding of 18th International Conference on Data Engineering. IEEE Press, New York, pp 673–684

Whitehead RL, Schiavetti N, Whitehead BH, Metz DE (1997) Effect of sign task on speech timing in simultaneous communication. J Commun Disord 30:439–454
Zacks JM, Tversky B (2001) Event structure in perception and conception. Psychol Bull 127:3–21
Zacks JM, Tversky B, Iyer G (2001) Perceiving, remembering, and communicating structure in events. J Exp Psychol Gen 130:29–58