Quantifying Expert and Impaired Imitative Learning

Brian J. Gold
Volcan Center for Complex Systems, Brandeis University

Introduction

Imitation is a prerequisite for the acquisition of language and other skilled behaviors. One way in which humans learn new skills is through observation and subsequent copying of the observed behavior, which Byrne and Russon (1998) refer to as 'action level' imitation. Action level imitation is differentiated from program level imitation in that the latter is broader and, whereas action level imitation contains linear, sequential acts, program level imitation is not necessarily linear. That is, program level imitation can contain multiple action level imitations or repetitions of the same action level imitation. An example of program level imitation would be a toddler copying how his parents eat, which is comprised of many action levels, including picking up the food, and putting it in his mouth. Repeating those two actions results in the program level imitation "eating". Action observation activates cortical motor representations, with the cortical bonds between the premotor and primary motor areas strengthened through repeated experience (Heyes, Bird, Johnson, & Haggard, 2005). Studying motor imitation, or motoric reproductions of observed actions, can help us understand the neural and cognitive mechanisms that make imitation possible and useful. Here we present three studies designed to take a quantitative look at action level imitation, using a novel methodology to study expert and impaired imitation. The first study demonstrates the usefulness of our approach, while the second and third examine the effect of expertise and psychiatric disorders, respectively, on imitation.

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

Supported by NSF grant SBE-0354378. email:bjgold@brandeis.edu
A large part of the methodology and Study 1 are taken directly from (B. J. Gold, Pomplun, Rice, & Sekuler, 2008) Thank you to Dr. Jim Morris, Kimberlee Bachman, and Amanda Hennesch for their help in identifying the ASL gestures, and to Dr. Robert Hoffmeister and Dr. Marlon Kuntze for advice regarding Study 2. Study 3 was performed in collaboration with Dr. Sohee Park, Dr. Natasha Matthews, and Crystal Gibson at Vanderbilt University.
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, Kollios, and Gunopulos (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 scale. 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 could elide theoretically-important characteristics of the imitation, including sequencing and serial-order effects (e.g., Lashley, 1951; Agam, Bullock, & Sekuler, 2005; Agam & Sekuler, 2007).

Imitation fidelity can therefore be determined by using the spatiotemporal properties of the imitation to break it down into multiple basic components. It has been shown that human observers break down imitations into component parts (Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001), and thus any algorithm developed to quantify imitation must do the same if it is to be comparable to human performance. Once an imitation has been decomposed to its component parts the spatial and temporal differences between the imitation and the model can then be studied.

There are multiple ways in which an imitation could deviate temporally from the original model. The imitation could differ from the model in the total amount of time taken to produce, the amount of time taken to complete each component in the model, and the amount of time taken to transition from one component of the model to the next. Data from how these temporal properties change over time gives us information as to the cognitive processing of the stimulus and the rate of learning, such that in many cases subjects perform subsequent repetitions of a model faster than initial repetitions (e.g., Agam et al., 2005; Agam, Galperin, Gold, & Sekuler, 2007). In addition, any lag before participants start their imitation could be correlated with the initial processing and motor planning as well as memory consolidation of the representation of the model (e.g., Hauptmann & Karni, 2002).

There are three different kinds of errors that can be made: gesture errors, sequence errors, and unmatched errors. These are described later, but reflect different types of processing errors. Gesture errors refer to differences in a single, specific component, such as flexing incorrect fingers when trying to make a specific hand gesture. Sequence errors refer to the incorrect ordering of the components in the imitation, and are in some studies referred to as temporal errors (for example Weiss et al., 2001). Unmatched errors refer to extra components that are included in the imitation or those that are left out of the imitation but were a part of the model.

Thus, by scoring an imitation as simply correct or incorrect, one may omit crucial information. Specifically, this information can help us to understand why the imitation deviates from the model. We can use this to determine the best ways to teach for different
The algorithm presented here is the first step in achieving this. This algorithm can segment a trajectory into its components in both spatial and temporal space, allowing us to make the various spatiotemporal comparisons to a second trajectory. In addition, the algorithm is an extensible algorithm, so that while the studies discussed here use only data from the fingers of one hand, with minor revisions the algorithm can be adapted to compare more complex imitations, such as whole body imitations.

General Methodology

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 contained a specific number of component gestures, to compare the imitation sequences to the model sequences, the algorithm first needs to divide both the imitation and model into their individual components and then compare these corresponding components. Toward this end, the algorithm resamples the data to achieve a constant sampling rate of 50 Hz for all data values. 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 studies presented here, the window was set to 140 ms, or 7 samples - from $(t − 3)$ to $(t + 3)$:

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

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 the flexion and extension of all five fingers relative to the palm during the time interval when the velocity drops below 10% of the peak velocity (the maximum velocity of all flexion values across the entire sample) and remains below that level for at least 100 ms. The duration threshold of 100 ms was empirically determined to yield most plausible results for the present type of motion sequences. The time period between two consecutive component gestures, that is the time in which some or all of the digits are in motion, is the transitional time, and movements during that time are termed transitional movements. Strictly speaking, 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 1 shows the results of the segmentation process for one trial, as well as the pictorial 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. 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 can produce, 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. An example of a gesture-level error is the flexion swapping between two digits. For example, if the model 12–124 were reproduced as 12–125, then the subject simply extended the little finger instead of the ring finger in the second gesture.

Note that in the above case, 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 2 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 & Estes, 1977; Lewandowsky & 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 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 an imitated gesture as a function of that gesture’s serial position in the model sequence. Whereas sequence-level analysis is performed on any length sequence, serial position analysis is performed on only those sequences in which the number of gestures reproduced matches the number of
gestures in the model sequence. Each item in the model sequence is then compared to each in the reproduction. For the serial position analysis, an item is correct if it matches the corresponding item in the model sequence in both flexion and order.

**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 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 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 spatial errors (swapping, inserting, and omitting flexions), which are determined based on the difference in flexions between the previous gesture and the current gesture. 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.

Materials

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. For the initial study only, the size of the stimulus model hand was 8.6° visual angle at its longest point, from the wrist to the tip of the middle finger, while the widths of the models wrist and palm, with all digits extended, were 2° and 5.7° visual angle, respectively. For all other studies, the visual angles measured 9.4°, 2.7°, and 10.8°, respectively.

Allowing all five digits of one hand to be flexed or extended produces a set of $2^5 = 32$ hand gestures. Because we knew that all 32 gestures would not be equally easy for subjects to reproduce (Schieber & 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. 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 to produce 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, seen in Figure 3A, used in Studies 1 and 3. Gestures used in Study 2 are shown in Figure 3B.
EXPERT AND IMPAIRED IMITATION

(a) Gestures for Studies 1 and 3

(b) Gestures for Study 2

Figure 3. A: The gestures used for the sequences in Studies 1 and 3. B: The gestures used for the sequences in Study 2. Letters to the left of the white line in B closely resemble letters in ASL and were used to construct the signed-sequences. Letters to the right are unrelated to ASL letters and signs and were used to construct the nonsense-sequences.

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 averages 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.
Study 1: A test of methodology

Intro & Motivation

In the initial application of our analytic method we examine 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 Figure 3A, 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, these stimuli were attractive because of their kinship to gestures in American Sign Language (ASL), which is studied in Study 2. Here, we present some experimental data that demonstrates our approach and shows that our method can successfully generate a useful multivariate description of various types of errors made during imitation of gesture sequences, and in characterizing practice-based improvements in imitation.

Methods

Subjects

Eight paid subjects were recruited from the Brandeis University community, and were between 19 and 29 years old ($M = 21.25, SD = 3.05$). All subjects reported no prior experience with American Sign Language (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 study 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).

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 study itself. With these predefined hold and transition times for gestures, our sequence-generating software used tweening\footnote{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.} to interpolate 42 frames between gestures in the sequence, blending successive static gestures with seamless transitions, as seen in Figure 4B. 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 six-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 of the one-transition condition, a total of five digits can change from the first gesture to the sixth; with stimuli from the two-transition condition, a total of ten digits can 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. Note that the number of digits that changed from the initial open hand to the first gesture and from the sixth gesture to the final open hand were not constrained to follow the one-transition and two-transition requirements; the transition requirements applied only from the first through sixth gestures.

Procedure

Subjects viewed the stimuli seated at a table, their right elbow on a wrist rest with their forearm and digits extended straight up, with their palm facing them, 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 stimulus transform. For each 6-item sequence, stimulus presentation took 11.5 seconds. After the presentation completed, the screen cleared, and subjects waited for a tone (two seconds after the stimulus had completed) before beginning their imitation. Subjects were instructed to accurately imitate as many gestures in the sequence as possible, and to try to attempt to reproduce them in the correct order. Subjects were allowed 14 seconds 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
Figure 4.  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.

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-second 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 Figure 3, was displayed for 1 second. Two seconds after the gesture disappeared from the display screen, the subject heard a tone indicating that they were
Figure 5. Figure depicts the presentation of one trial. Subjects start each trial by pressing the space bar, after which the model stimulus appears on the screen. The model stimulus is a sequence of 6 gestures, which takes 11.5 seconds to complete. 1 second after model stimulus presentation a tone sounds, which instructs subjects to begin their imitation. Subjects have a total of 14 seconds to complete their imitation, during which time they view a blank screen. At the end of the 14 seconds, subjects are instructed to press the space bar to continue to the next trial.

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 two seconds (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 second, with a transitional time of 500 ms. Thus the total stimulus presentation time for these two-gesture sequences was 5.5 seconds. As in the first familiarization phase, two seconds 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 seconds 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 six-item gesture sequences, chosen from the set of 16 sequences. Each sequence began and ended with an open hand, and included tweened frames between successive gestures. Each gesture was held for 1 second, with a 500 ms transitional time between gestures. The eight model sequences were shown and imitated in
massed fashion; that is, each model sequence was viewed and imitated 10 times in a row 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.

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. 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.

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, premovement 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 to 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.

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, \( F_{1,7} = 1.489, p = .262 \). Repetition of a model significantly influenced the number of gestures reproduced, with that number increasing systematically from the first to the tenth presentation \( F_{6.777,47.437} = 9.909, p < .001 \). This effect can be seen in Figure 6A. There is also a significant interaction between transition condition and repetition \( F_{9,63} = 5.910, p < .001 \), 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.
Figure 6.  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.
Errors in Imitation

The (normalized) total number of errors decreased with repetitions, demonstrating an improvement in performance with practice ($F_{4.861.29.165} = 4.853, p < .001$). 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, ($F_{5.642.33.849} = 3.945, p < .01$). 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. Figures 6B and C show 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 ($F_{6.635.39.811} = 6.093, p < .001$). In addition, subjects make approximately twice as many gesture level errors in the two-transition condition compared to the one-transition condition ($F_{1.6} = 41.792, p = .001$). 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 ($F_{3.973.23.836} = 4.278, p = .010$). The same pattern is observed for the unmatched gestures, with improvements occurring with repetition ($F_{5.634.33.806} = 6.559, p < .001$). 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 ($F_{6.385.38.309} = 3.554, p < .01$).

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 serial position and digit flexion? To calculate serial position we compared each item in the model gesture to each corresponding item in the reproduction for those reproductions in which the subject reproduced the correct number of gestures, six in this case. A correct match was assigned a value of 1 and an incorrect match was assigned a value of 0. We find that fewer errors are made with repetition ($F_{5.556.38.889} = 4.828, p = .001$) and also with serial position ($F_{3.125.21.874} = 19.284, p < .001$), such that fewer mistakes are made in earlier gestures than later gestures in a sequence. This primacy effect is significantly larger for the one-transition condition compared to the two-transition condition (Figure 7A and B, respectively), shown by a significant interaction between transition condition and serial position ($F_{4.291.30.035} = 4.493, p < .01$).

Temporal Analysis of Imitation

Premovement 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 premovement latency from repetition one (Mean=1200.938, SD = 457.304) through to repetition ten (Mean=798.438, SD=268.301), indicating that they take significantly less time to prepare their response as they become more familiar with the stimulus ($F_{0.54} = 3.243, p < .01$).
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 number of gestures reproduced. We observe a decrease in movement time as repetition number increases from repetition one ($M=2149.734$, $SD=460.551$) to repetition ten ($M=1645.175$, $SD=334.742$); this effect is statistically significant ($F_{6,164,36.981}=4.089, p < .01$). In addition, subjects take longer to perform the two-transition stimuli ($M=1987.395$, $SD=376.572$) than the one-transition stimuli ($M=1555.560$, $SD=323.188$); this main effect of transition condition is also significant ($F_{1,6}=6.533, p < .05$).

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 ($F_{1,7}=12.088, p = .01$).

Discussion

The current experiment presented and tested 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 & 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.

Here, we have presented data that illustrates 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 visuo-spatial memory. In the one-transition condition, we found both a primacy effect and a trend towards a one-item recency effect, as there is a significant decrease in proportion correct from the first to the second serial position but no significant decrease in proportion correct from the penultimate to ultimate serial positions. These findings are similar to findings from other studies of short term memory, using either verbal materials (Lee & Estes, 1977; Lewandowsky & Murdock, 1989) or non-gestural motor imitation (Agam et al., 2005, 2007).

In addition, we demonstrated that our algorithm could identify chronometric properties of imitation, with longer pre-motor latencies observed during subjects’ early imitations compared to their later imitations. This presumably reflects a decrease in cognitive effort required to encode and recall 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. However, the current study does not fully dissociate serial memory from motor imitation. It is possible that the temporal changes we see here are a result of synaptic plasticity in the hand representation area of the primary motor cortex over the time course of 10 repetitions of the same sequence, and thus reflect a decrease in the required motor effort rather than a decrease in the required cognitive effort. As a step towards separating memory from motor imitation, it would be beneficial to conduct a study in which subjects verbally reproduce the sequences from memory.

Our algorithm also recovered performance variations that arose 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, and the two conditions converging on later repetitions when subjects have had more practice at the sequences.

The algorithm demonstrated here allows us to compare an imitation of any multi-dimensional 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. The algorithm tested here can be used to examine differences in performance in motor learning between various populations, which we test in Studies 2 and 3.
Study 2: The effect of expertise and complexity on imitation accuracy and timing

Study 1 found a relationship between stimulus complexity and subjects’ performance when reproducing that stimulus, defining stimulus complexity by the number of changes in digit flexion and extension across all gestures in that sequence. That study demonstrated that the complexity of a sequence influenced the ease with which subjects reproduced the sequence, both in terms of spatial errors and the temporal properties of the imitation. There were approximately twice as many gesture-level errors made in subjects’ reproductions of the more complex (two-transition) stimuli compared to the less complex (one-transition) sequences. Additionally, Study 1 found that subjects imitated less complex sequences faster than they did more complex sequences, both in terms of overall movement time and in terms of the time taken to transition between successive pairs of gestures. Building on this, we hypothesize that sequences with fewer digit changes would result in fewer spatial errors and take less time to imitate than sequences with more digit changes.

An advantage of the methodology developed in Study 1 is that it affords the possibility of examining the imitation performance and errors produced by individuals with expertise in some particular movement-related domain. Specifically we can examine the effects of stimulus complexity on expertise. If expertise in a movement-related domain results in better performance and faster learning on novel, related tasks, then it would be valuable to understand the differences between experts and non-experts. Through these differences we may be able to find a way in which the non-experts can achieve the same level of learning as the experts without undergoing the rigorous training of the experts.

The stimuli in Study 1 were hand gestures, and as previously noted, some of those gestures very closely approximate letters and digits from the ASL fingerspelling alphabet (see Figure 3A). As a result, when imitating sequences of ASL gestures, individuals who have studied ASL may have an advantage over those who have not studied ASL (for example, Whitehead, Schiavetti, Whitehead, & Metz, 1997; Jerde, Soechting, & Flanders, 2003). This conjecture is supported by the demonstration that in a short-term temporal recall task both deaf and hearing subjects can exploit phonetic coding of seen gestures (Hanson, 1990). In addition, Wilson and Emmorey (1997) suggested the possibility of a language-based rehearsal loop, similar to the Baddeley and Hitch (1974) phonological loop for speech, providing support with an experiment in which pictures were translated into an ASL-based code for memory.

Supporting the hypothesis that the ability to label gestures as letters when viewing sequences of hand gestures improves imitation performance when compared to no labeling, Frencham, Fox, and Maybery (2004) found enhanced serial recall when verbal labeling was encouraged in a memory task involving hand movements as stimuli. Tanaka, Inui, Iwaki, Konishi, and Nakai (2001) found a difference in activation of the supramarginal gyrus when Japanese participants imitated hand gestures that had symbolic meaning in Japanese culture compared to the imitation of meaningless hand gestures. The supramarginal gyrus connects language and speech production (Geschwind, 1965) and is also involved in integrating audition with movement, as shown in sign language (Emmorey, Mehta, & Grabowski, 2007) and movement preparation and imagery (Krams, Rushworth, Deiber, Frackowiak, & Passingham, 1998; Hanakawa, Dimyan, & Hallett, 2008). It is therefore likely that signers
will perform better on sequences that involve meaningful gestures compared to meaningless gestures. Meaningful gestures are hand configurations that have some distinct definition for a subject, whereas meaningless gestures are defined as gestures which have no significance for a subject. In the present study, in addition to studying stimulus complexity, we explore whether basic knowledge of ASL fingerspelling is sufficient to give signers an advantage over nonsigners in a short-term memory-based imitation task. Further, we examine whether the advantage of signers over nonsigners is global for hand gestures or is limited to only those hand gestures with which they have prior experience. We hypothesize that any advantage signers have over nonsigners will be limited solely to those sequences that contain gestures that are meaningful to the subject; that is, expertise will play a role only in the sequences directly involving the ASL gestures and only for those subjects who have prior experience with ASL.

Methods

Subjects

Twenty hearing, native English speakers participated in this experiment (Age: $M = 21.350, SD = 3.117$). Using a task designed to measure ASL fingerspelling competency (described below), subjects were assigned to one of two groups, depending on their competency with ASL; those with experience with ASL were placed into the ASL-experienced group, and those without were in the ASL-naive group. Ten subjects comprised the former group (Age: $M = 21.700, SD = 3.268$), while 10 comprised the latter (Age: $M = 21.000, SD = 3.091$). All subjects had normal or corrected-to-normal vision. Additionally, all subjects were classified as right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). In accordance with the principles of the Declaration of Helsinki, all subjects provided written informed consent. The experimental protocol had been approved by Brandeis University’s Committee for the Protection of Human Subjects. Subjects were compensated for their participation, either monetarily or through credits for an introductory psychology course.

Stimulus Construction

Stimuli were constructed using the same Matlab program referenced in the Stimulus Construction section of Study 1. However, instead of the 16 gestures used in Studies 1 and 3, the present study used the 14 gestures shown in Figure 3B, all of which were biomechanically possible for each subject to reproduce. Of the 14 gestures, half closely resemble letters in the ASL fingerspelling alphabet, whereas the other half do not resemble any ASL gesture or sign. It has been shown that numbers are not comparable to other stimuli in short term memory, activating areas that are not active during letter stimuli (Knops, Nuerk, Fimm, Vohn, & Willmes, 2006). As a result, gestures that closely resembled numbers in the ASL lexicon, or that resembled commonplace slang or symbols were not used in this study. Three volunteers, fluent in ASL, viewed each of our 14 gestures and individually verified that each gesture was correctly classified as closely resembling an ASL letter or as having no resemblance to any sign in ASL.

As in the first study, experimental sequences all began and ended with an open hand. Additionally, sequences were constructed such that for every possible pair of gestures, no successive pair of gestures could appear in more than one sequence. Because of the limited
number of gestures available for use in this study, experimental sequences contained only five gestures between the open hand at start and end of a sequence.

To study the influence of expertise, two distinct types of sequences were used, one which contained only gestures that closely resembled letters in the ASL fingerspelling alphabet (signed-sequences), and one which contained only meaningless, non-ASL gestures (nonsense-sequences). A further constraint on the signed-sequences was that the corresponding letters involved could not spell an English word inside the sequence. Ten sequences of each type were created. As in Study 1, each component gesture was held for 1000 ms, with a 500 ms transition between each pair of gestures in the sequence.

To look at the effect of stimulus complexity on performance, all 20 sequences (regardless of ASL composition) were divided into three categories based on the number of changes in digit flexion and extension in each sequence, lowDC (7-9 digit changes), midDC (10-11 digit changes), and highDC (12-13 digit changes), with 6, 7, and 7 sequences in each respective category. For example, the sequence 5-345-2345 would be an example of a highDC sequence, with a total of 13 digit changes comprising the sequence. An example of a lowDC sequence would be 1245-1345-12345, with 7 changes. The nonsense-sequences comprise the bulk of the lowDC and midDC categories. Six of the seven sequences in the highDC category are signed-sequences.

Procedure

Imitation Task. After calibrating the data glove, subjects twice viewed and imitated each of the 14 static gestures, presented one at a time. The calibration routine and timing for the static gestures was the same as in the previous study. Subjects then viewed and imitated three different two-item sequences whose component gestures were not among the 14 gestures in figure 3B. Finally, subjects viewed and imitated 16 different 5-item sequences, split equally from the set of signed-sequences and nonsense-sequences. Each sequence was shown and imitated five times in massed fashion before the next sequence was displayed. The order of sequences to be displayed was determined in a randomized block fashion between signed-sequences and nonsense-sequences, and was counterbalanced across all subjects.

ASL Proficiency Task. After completing the imitation task just described, all subjects underwent a test of ASL fingerspelling ability. This proficiency test was given after the imitation portion of the experiment so as not to prime the subjects for the purpose of the experiment. In addition, only the ASL-like gestures were used during this test; having this test before the imitation task would bias the types of sequences, as the two types of gestures would have been seen an unequal amount of time prior to the actual imitation data collection.

During the proficiency test, subjects viewed one of the seven gestures that closely resembled an ASL fingerspelling letter on one side of the screen, and a letter of the English alphabet on the other. Subjects were instructed to respond as quickly as possible with a keypress as to whether or not the gesture and letter had the same meaning. Each letter/gesture pair disappeared from the screen after 1s, but subjects were allowed an additional 500 msec to make their response. Failure to respond was scored as an incorrect response.
One third of the trials were target trials, that is, the letter of the English alphabet corresponded to the ASL fingerspelling gesture (targets). One third of the trials were ASL lures in which the letter corresponded to an ASL fingerspelling gesture shown during the imitation task, but not to the particular ASL gesture that was displayed on the screen (ASL-lures). The remaining one-third of the trials comprised pure lures: the letter shown did not correspond to any of the ASL fingerspelling gestures used in the imitation task (pure-lures). For each gesture, only three letters were used: one match, one ASL lure, and one pure lure. A letter used during an ASL-lure trial for a specific gesture was not reused in a different ASL-lure trial; the same held true for the pure-lures trials. Response keys to signify a match or non-match were counterbalanced across subjects, and gestures and letters were displayed equally often on either side of the screen. Each gesture-letter pair was shown 12 times to each subject, producing enough samples for a reliable performance measure.

Results

Stimulus Complexity

To look at the effect of stimulus complexity on performance, we evaluated all subjects’ imitations on the three different complexity categories. As mentioned earlier, categories were defined by the number of changes in digit flexion and extension in each sequence (lowDC, midDC, and highDC). For this initial analysis we did not separate subjects on the basis of their ASL knowledge.

There were fewer total errors (normalized by the number of gestures produced) produced when subjects imitated the LowDC sequences than the other sequences, $F_{2,38} = 6.241, p = .005$. Similarly there were fewer gesture-level and sequence-level errors when subjects imitated the LowDC sequences (Gesture-level: $F_{2,38} = 20.936, p < .001$, Sequence-level: $F_{1,520,.28,889} = 6.123, p = .010$). However, subjects also make more unmatched gestures for the LowDC sequences, $F_{1,757,.33,382} = 3.447, p = .049$. Means and standard deviations are shown in Table 1. As hypothesized, subjects make fewer gesture and sequence errors on the sequences with fewer digit changes.

<table>
<thead>
<tr>
<th>(norm) Spatial Error Category</th>
<th>LowDC</th>
<th>MidDC</th>
<th>HighDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Errors</td>
<td>0.854 ± 0.118</td>
<td>1.007 ± 0.083</td>
<td>1.071 ± 0.108</td>
</tr>
<tr>
<td>Gesture Errors</td>
<td>0.373 ± 0.048</td>
<td>0.540 ± 0.060</td>
<td>0.587 ± 0.060</td>
</tr>
<tr>
<td>Sequence Errors</td>
<td>0.092 ± 0.017</td>
<td>0.164 ± 0.018</td>
<td>0.149 ± 0.019</td>
</tr>
<tr>
<td>Unmatched Gestures</td>
<td>0.390 ± 0.067</td>
<td>0.303 ± 0.045</td>
<td>0.338 ± 0.051</td>
</tr>
</tbody>
</table>

Table 1: Means for spatial error categories by stimulus complexity for Study 2, normalized by number of gestures produced in imitation.

Additionally, subjects were faster to transition between subsequent gestures in a sequence if that sequence contained fewer digit changes (LowDC), $F_{2,38} = 23.342, p < .001$. The shortest average transition time, during the imitation of the LowDC sequences, at 660.205(24.906) ms, was still slower than the 500 ms transition time between gestures in the model sequences. This result is not simply driven by the physical time taken to flex or
extend a finger; due to biomechanical constraints of the hand (Schieber & Santello, 2004), pairs of gestures with the same number of changes in digit flexion will have different transition times (for example, a transition between the gesture pair 1-2, due to the independence of the fingers, will take less time than a transition between highly coupled fingers, such as in the gesture pair 45-5). Similarly, if the digits that change between a pair of gestures are independent, the transition between gestures would be faster than between a pair of gestures where fewer digits change, if the digits are more dependent (such as the case comparing 12345-23 versus 45-5). Thus, a faster transition time for the LowDC sequences reflects the relative ease in imitating those sequences compared to the HighDC sequences, rather than the time taken to simply physically flex or extend the digits in a sequence.

Average hold times tell a different story however, with subjects holding the gestures in sequences containing the most digit-changes for the shortest amount of time, $F_{2,38} = 4.082, p = .025$.

Reproduction becomes easier with repetition

As in Study 1, subjects produce more gestures as sequences are repeated, $F_{2,163.41 \text{.}097} = 16.676, p < .001$. The normalized movement time (NMT) is a measure of the speed with which subjects perform their imitation, such that movements that are more well-learned are performed more quickly than comparable, but novel, movements. Thus, the more times subjects reproduce a sequence, the more gestures should be made, as seen above, and the more quickly the sequence should be imitated. The NMT is calculated as the total time taken by the subject to complete the imitation, excluding the initial and final open hands, divided by the number of gestures produced. As expected, the NMT decreases with repetition, $F_{1,856.35 \text{.}270} = 6.562, p = .005$, such that subjects complete their imitations faster with repetition. Subjects also transition from one gesture to the next more quickly with repetition, $F_{2,897.55 \text{.}052} = 13.599, p < .001$, also suggesting that subjects are learning the component gestures of a sequence they more they view and imitate a sequence.

In addition to performing their reproductions more quickly with repetition, subjects also make fewer normalized errors in each error category with repetition (Total errors: $F_{2,652.50 \text{.}384} = 46.131, p < .001$; Gesture: $F_{2,260.42 \text{.}934} = 7.791, p = .001$; Sequence: $F_{3,232.61 \text{.}413} = 31.577, p < .001$; Unmatched gestures: $F_{1,452.27 \text{.}583} = 11.679, p = .001$). Subjects also make fewer sequence errors on the first repetition of the LowDC sequences compared to the other sequences. This difference in first repetition drives a significant interaction of stimulus complexity and repetition for the sequence level errors, $F_{6,620.125 \text{.}775} = 2.757, p = .012$.

The effect of expertise on stimulus complexity

The above analyses do not take expertise into consideration when looking at stimulus complexity. Using the ASL proficiency task, subjects were separated into groups based on their level of knowledge of ASL. This allows us to examine the effect of expertise on stimulus complexity.

**ASL Proficiency** The ASL proficiency task was used to group subjects by their comprehension of ASL, specifically ASL fingerspelling. In a pre-study questionnaire, subjects self-reported their knowledge of ASL, but this self-report was not always predictive of their
actual performance on the proficiency task. Using subjects’ reaction times, accuracy, and d-prime statistic, subjects were placed into one of two groups, ASL-experienced and ASL-naive. The two groups were not constrained a priori to be equal size, but that did occur. D-prime was calculated as the remainder of the normalized false alarm rate subtracted from the normalized hit rate. The false alarm rate was defined as the proportion of all lure (both pure-lure and ASL-lure) trials misidentified by the subject as a target. The hit rate was defined as the proportion of target trials that were answered correctly.

There was a clear cutoff between the groups by d-prime, with those subjects in the ASL-experienced group recording a much higher hit rate and fewer false alarms compared to subjects who had little to no experience with ASL, (D-prime: ASL-experienced: $M = .284, SD = .024$; ASL-naive: $M = .106, SD = .049$). This difference was significant, $F_{1,16} = 101.961, p < .001$, yielding 10 subjects in the ASL-experienced group and 10 in the ASL-naive group. There was no significant difference in response time, but there was a significant difference in accuracy, with subjects placed in the ASL-experienced group producing over 50% more correct responses than those placed in the ASL-naive group, $F_{1,16} = 63.981, p < .001$. Two subjects were excluded from ASL proficiency task due to computer problems, but as both self-reported having no prior knowledge of ASL, and neither was able to correctly identify the English meanings of any gestures during the imitation portion of the experiment, their data from the imitation task was still included in the analyses below.

**Expertise** Separating the subjects into two groups based on experience yields significant interactions that differentiate experience from sequence complexity. Calculating the normalized movement time (NMT), ASL-experienced subjects perform their fastest imitations on HighDC sequences, while the ASL-naive subjects are fastest on the LowDC sequences, as shown in Figure 8A, $F_{2,36} = 4.269, p = .022$. However, the NMT for all conditions and groups is still slower than the 1.6 s NMT for the model sequences. As previously mentioned, of the 7 sequences in the HighDC group, 6 are comprised solely of ASL gestures (signed-sequences). Thus the NMT was faster for the ASL-experienced subjects only on the signed-sequences, which are composed of ASL gestures with which they had prior experience. ASL-experienced subjects’ NMT on the nonsense-sequences were similar to that of the ASL-naive subjects’ for both the signed- and nonsense-sequences, this is displayed in Figure 8B. Thus the performance of the ASL-experienced subjects on the signed-sequences drives a significant interaction between sequence type and subject group, $F_{1,18} = 11.556, p = .003$.

ASL-experienced subjects also produced fewer normalized total errors and unmatched gestures compared to ASL-naives when initially imitating a sequence; these differences disappear after the first repetition (Total errors: $F_{3.099,55.774} = 55.657, p < .001$; Unmatched gestures: $F_{1.687,30.364} = 4.602, p = .023$).

Finally, in examining at the average hold time for a reproduced gesture in a sequence, there is a significant interaction between stimulus complexity and subject group, $F_{2,36} = 4.317, p = .021$. This difference is solely due to the expert groups’ performance on the sequences with the most digit changes; as shown in Figure 9A, subjects with prior experience hold the sequences with the most digit changes almost 300 ms less than they hold any other sequence and less than the novice group holds any sequence. As the HighDC group was composed of mostly sequences containing ASL gestures (signed-
Figure 8. A. NMT for each subject group when dividing the sequences into number of digit changes in flexion and extension. Note that ASL-experienced subjects are faster in imitating the HighDC sequences than the LowDC sequences, while the ASL-naive subjects are faster when imitating the LowDC sequences than the HighDC sequences. B. NMT for each subject group and sequence type. Note the much shorter time for the signed-sequences when imitated by ASL-experienced subjects compared to any of the other conditions. For both A and B, the black horizontal line indicates the normalized movement time of the model sequence.

sequences), the hold time was shorter for the ASL-experienced subjects only on those sequences containing gestures with which they had prior experience, the signed-sequences. This is confirmed by a significant interaction between subject group and sequence type (signed and nonsense), $F_{1,18} = 19.078, p < .001$. Figure 9B shows that the ASL-experienced subjects held the signed-sequences for less time than they did the nonsense-sequences ($M = 760.759, SD = 66.148$ and $M = 1026.494, SD = 74.337$, respectively). There is no difference between the sequence types for the ASL-naive group.

Discussion

The present study showed that, as hypothesized, less complex sequences were easier to imitate than more complex ones. Subjects made fewer spatial errors and spent less time reproducing the sequences and transitioning between gestures in the less complex stimuli than they did imitating the more complex stimuli. Additionally, the present study demonstrated that experience modulates the effect of stimulus complexity. Subjects who had prior experience with stimulus components spent less time reproducing the stimulus and less time paused on each individual component of the stimulus, but only when the individual components were directly related to the subjects' prior experience. In this case, these results applied only to the ASL-experienced subjects on the signed-sequences, as all ASL-experienced subjects had prior experience with ASL, specifically fingerspelling.

ASL-experienced subjects had faster NMTs in completing their reproductions in the signed-sequences than the nonsense-sequences. Additionally, ASL-experienced subjects held...
(a) Average Hold Time by subject group and number of changes in finger flexion
(b) Average Hold Time by subject group and sequence type (meaningful vs. meaningless)

Figure 9. A. Average hold time for each subject group when dividing the sequences into number of changes in finger flexion. Note the much shorter hold time for the HighDC sequences when imitated by the ASL-experienced subjects. B. Average hold time for each subject group and sequence type. Note the much shorter hold time for the signed-sequences when imitated by ASL-experienced subjects compared to any of the other conditions. For both A and B, the black horizontal line indicates the hold time for each gesture in the model sequence.

Each component gesture for less time when reproducing the signed-sequences. There was no difference between the sequence types for the ASL-naive group, with their times similar to those of the nonsense-sequences of the ASL-experienced group, as shown in Figure 9B. Our results here are in line with previous studies of expertise, demonstrating temporal advantages afforded experts in their field of expertise (Halpern & Wai, 2007; Overney, Blanke, & Herzog, 2008).

The present study shows that stimulus complexity and expertise both affect imitation accuracy and timing. Future studies should attempt to separate expertise from complexity. Manual or articulatory suppression while viewing the sequences could cause the ASL-experienced subjects to lose their temporal advantage for the signed-sequences over both nonsigners and the nonsense-sequences, especially as a significant correlation between articulatory and manual suppression has been documented (Xue, Aron, & Poldrack, 2008). Additionally, as mentioned above, our subjects were not fluent in ASL. It would be beneficial to have both native bilingual subjects as well as deaf native signers participate in this experiment to compare their performance to that of hearing nonsigners, expanding our study of stimulus complexity and expertise.
Study 3: Differences in imitation between psychiatric populations

Study 2 showed that expertise with the gestures in some stimulus category (expertise) afforded subjects an advantage in learning via imitation. However, there are also subject groups for which learning and imitation is difficult. Characterizing and evaluating the different imitation strategies is important in the evaluation of treatments to neurological diseases which may result in impaired imitation, such as in schizophrenia. Schizophrenic patients have been shown to have marked deficits on many memory tasks (for example, Knight, Elliott, & Freedman, 1985; Fleming et al., 1997; Leiderman & Strejilevich, 2004; Park & Holzman, 1992). Other studies have shown that schizophrenics are significantly impaired compared to control subjects on spatial sequence learning tasks (Pedersen et al., 2008; Fraser, Park, Clark, Yohanna, & Houk, 2004; Foerde et al., 2008), and on verbal working memory tasks (Elvevag, Fisher, & Goldberg, 2003; Raalens, Ramsey, Jansma, Jager, & Kahn, 2008; Silver & Goodman, in press). Further, while control subjects show typical learning upon repeated exposure to stimuli, schizophrenics show slower learning (Pedersen et al., 2008; Foerde et al., 2008).

However, the vast majority of these tasks have used either verbal or spatial stimuli. Very little has been done to study the sequence learning by schizophrenics with a visuomotor task, with the relevant studies generally incorporating a motor component into a spatial memory task so that the subject is required to point to or touch a target (Foerde et al., 2008; Fraser et al., 2004; Leiderman & Strejilevich, 2004; Tabares-Seisdedos et al., in press). Even studies that use visuomotor memory tasks limit themselves to finger tapping (Silver & Goodman, in press) or simple hand movement to a target (Park & Holzman, 1992). The present study seeks to expand upon these studies, through a task in which gesture sequences are imitated and learned. The methodology used in Studies 1 and 2 (B. J. Gold et al., 2008) allows us to look not only at performance changes over repeated presentations and imitations of a sequence, but also look at the temporal differences between schizophrenic patients and controls.

Methods

Subjects

Thirty-six hearing, native English speakers participated in this experiment, and were divided into three groups: schizophrenic patients (SZ), patients with bipolar disorder (BP), and normal controls (CO). Subjects in the control group were age-, IQ-, and education-matched with the schizophrenic group (see Table 2). All subjects had normal or corrected-to-normal vision and were classified as right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). No subject reported having prior experience with American Sign Language (ASL); this exclusion criterion was important as some of our stimuli (Figure 3A) were similar to letters in the ASL 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 the Vanderbilt University Institutional Review Board and the Brandeis University Committee for the Protection of Human Subjects. Thirty subjects participated at Vanderbilt University, and 6 controls participated at Brandeis University. SZ were recruited from outpatient sources in Nashville,
Control subjects at Vanderbilt University were recruited by means of advertisements in the general community. Control subjects at Brandeis University were recruited through the general campus community.

Schizophrenia diagnosis was confirmed using the Structured Clinical Interview for DSM-IV Disorders (SCID). Potential subjects were excluded if they met criteria for alcohol or drug dependence, or had a history of head injury or neurological disorder. Clinical symptoms were obtained using the Scale of Assessment of Positive Symptoms (SAPS) and the Scale for the Assessment of Negative Symptoms (SANS). Controls were excluded if they met criteria for an axis 1 disorder on the SCID, or if they had a first degree blood relative diagnosed with schizophrenia or bipolar disorder. At the time of the study, patients with schizophrenia and bipolar disorder were taking medication for their disorders. IQ was determined using the Wechsler Abbreviated Scale of Intelligence (WASI). All subjects were compensated for their participation. Two subjects (one schizophrenic and one control) were removed from data analysis due to equipment failure.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Age</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenic (SZ)</td>
<td>15</td>
<td>38.933 (11.003)</td>
<td>15.533 (2.232)</td>
</tr>
<tr>
<td>Bipolar (BP)</td>
<td>6</td>
<td>40.333 (10.093)</td>
<td>14.500 (1.732)</td>
</tr>
<tr>
<td>Control (CO)</td>
<td>15</td>
<td>39.214 (9.333)</td>
<td>14.267 (2.374)</td>
</tr>
</tbody>
</table>

Table 2: Age (years) and Education (years) statistics for each subject group for Study 3

Stimulus Construction

Stimuli were constructed using the same Matlab program referenced earlier, using the same subset of possible hand gestures as in the initial study, shown in figure 3A. The number of gestures in a sequence was limited to either one or two. Shorter sequences were used in this study as Park, Matthews, and Gibson (2008) found that SZ were significantly less accurate than CO in imitating even two-item sequences of hand gestures. Each experimental sequence began and ended with an open hand. Sixteen one-item sequences were created, one for each gesture in Figure 3A. There were also 16 two-item sequences, created such that each gesture was used in exactly two sequences.

Procedure

There were two imitation conditions in the current study, memory and online. The memory condition was performed as in the previous studies, with subjects imitating each sequence after 2 s delay. In the online condition, subjects imitated each sequence concurrently with the stimulus display.

Familiarization Task. Each subject began the session by calibrating the data glove to fit his or her hand. After undergoing the calibration routine, subjects twice viewed and imitated each of the 16 static gestures online. Static gestures were displayed for 1 s, but subjects had an extra second to complete their reproduction so they would not run out of time. After completing the static online task, subjects viewed and imitated each one-item
sequence *online*. The one-item sequences were displayed three times each in massed fashion, with each sequence taking 4 s to display. Again, subjects were given an extra second to complete their imitation.

**Experimental Phase** After the practice trials, subjects performed two blocks of each imitation condition (*online* and *memory*) of the two-item sequences in the experimental phase. The order of the blocks was fully counterbalanced across subjects, and regardless of block, each sequence was displayed and imitated 8 times in massed fashion. Each gesture was seen exactly once in each task, such that no sequences overlapped between the *online* and *memory* tasks for an individual subject. All two-item sequences were 5.5 s in length. Subjects were given an extra second at the end to finish their imitations during the *online* task, while in the *memory* task, subjects had 10 s during which their response was recorded.

**Results**

**Statistical Analysis**

Dependent measures included all those detailed in Study 1. Each measure was subject to a repeated-measures ANOVA, with task (online or memory) and repetition (1 to 8) the within-subject variables, and subject group (schizophrenic, bipolar or control) as the between-subjects variable. Hunyh-Feldt corrections were applied in the case of the violation of sphericity assumptions. A significance threshold of 0.05 was used throughout.

**Number of Gestures Produced**

Unlike Studies 1 and 2, there was no significant effect of the number of gestures that the subjects reproduced. However, sequence models in Studies 1 and 2 contained 6 and 5 gestures respectively, and the current study contains only 2 gestures. It is therefore likely that the lack of significant findings as per the number of gestures produced in imitation is due to the low number of gestures that comprise each model in the current experiment, averaging 1.811 (.046) gestures produced during the first imitation of a sequence and 1.806 (.043) produced during the final imitation of a sequence.

**Errors in Imitation**

The normalized total number of errors decreased with repetition, indicating that with practice, performance improves ($F_{5,730,160,442} = 3.920, p = .001, \eta^2 = .140$). No other factor had a significant effect on the normalized total number of errors. Similarly, the normalized number of gesture errors, that is, flexion errors made during the reproduction of a specific gesture, decreased with repetition, showing that with practice, subjects make fewer finger flexion errors ($F_{5,337,149,450} = 5.995, p < .001, \eta^2 = .214$). There was a significant difference in the rate of improvement of gesture errors by subject group, as shown in Figure 10, with both SZ and CO showing a gradual decrease in gesture errors over repetition, while BP show no decrease, $F_{10,675,149,450} = 1.890, p = .047, \eta^2 = .135$ (SZ: $T_{13} = 2.730, p = .017$; CO: $T_{13} = 2.791, p = .015$; BP: $T_{1} = 1.728, p = .159$).

When subjects are imitating sequences in the *online* condition, it is highly unlikely they will make sequence errors, as they are imitating concurrently with the stimulus on the screen. As such, in looking at sequence errors, we look only at the *memory* condition. In
**Figure 10.** Gesture errors, or errors in finger flexion, for each of the three subject groups across repetition, normalized by the number of gestures made in imitation. Note that all groups decrease with repetition, but SZ improve the fastest, with the most improvement in the initial repetitions.

The memory condition, subjects make fewer sequence errors with practice, $F_{3.976,115.298} = 2.658, p = .037, \eta^2 = .085$. Subject group has no effect on sequence errors.

There is no significant effect for the unmatched gestures error category. Again, this probably reflects the low number of gestures in the sequences.

**Serial Order**

Serial position calculations were performed for the memory condition, for as noted regarding the sequence-level errors, it was highly unlikely subjects would make serial position errors in the online condition. The serial position calculations were performed as in Study 1, assigning a value of 1 to an item if the model gesture was reproduced correctly in both flexion and position; otherwise it was given a value of 0. Serial position accuracy improves with repetition ($F_{6.595,204.4387} = 4.089, p < .001, \eta^2 = .117$). Subjects are also more accurate on the first gesture in the sequence than the second, $F_{1.31} = 88.928, p < .001, \eta^2 = .742$. There is no significant difference in serial position by subject group.

**Temporal Analysis of Imitation**

In each sequence, the initial open hand was displayed on the screen for 1 s before transitioning to the first gesture. While subjects took about that long to begin in the memory condition ($M = 974.137, SD = 63.070$), they took significantly longer, about 500 ms longer before moving on to the first gesture in the online condition ($M = 1480.162, SD = 32.321; F_{1.28} = 64.052, p < .001$). Additionally, SZ take more time before beginning their imitation from memory than do the other two subject groups, as shown in Figure 11(a). Finally, though subjects speed up the time they hold the initial open hand for in the online condition, there is no change in repetition for the memory condition, $F_{7.196} = 2.824, p = .008, \eta^2 = .101$; this difference however is reflected in the large change between the first and second repetitions in the online conditions and can be seen in Figure 11(b).
EXPERT AND IMPAIRED IMITATION

(a) Premotor Time: Memory Condition

(b) Premotor Time: Condition by Repetition

Figure 11. A. Time taken by subjects before beginning the transition from the initial open hand to the first gesture (premotor planning time), in the memory condition. Initially, SZ take more time before beginning their imitation than the other two groups. Overall, CO take the least amount of time. B. Comparison of the two conditions (memory and online across repetitions for the premotor planning time, collapsing across subject groups. There is no change across repetition for the memory condition, but there is for the online condition.

In the model sequences, the transition between successive gestures took 500 ms. Though subjects are slower than the model transition in both the online and memory conditions, they take longer on the memory condition ($M = 720.294, SD = 37.159$) than the online condition ($M = 626.056, SD = 29.899$) by about 100 ms, $F_{1,28} = 8.774, p = .006, \eta^2 = .313$. Subjects do transition more quickly the more they imitate a sequence, but the difference is not significant, $F_{5.339,149.488} = 2.101, p = .064, \eta^2 = .0975$. The time subjects take to transition between gestures improves more rapidly with repetition in the online condition than the memory condition, which is pretty stable, $F_{6.700,187.586} = 2.180, p = .040, \eta^2 = .078$, as shown in Figure 12.

Discussion

Leiderman and Strejilevich (2004) claim that schizophrenic subjects poor performance relative to CO in both spatial and object visual working memory tasks means that visuospatial memory is impaired in SZ. Spindler, Sullivan, Menon, Lim, and Pfefferbaum (1997) conclude that poor performance is due to a combination of impaired spatial and object working memory. The task in the current study is a visuomotor task, as opposed to a simple object task. That we find no difference in spatial errors between the SZ and CO groups suggests that, in line with previous findings, there is no motor impairment in SZ (Foerde et al., 2008; Pedersen et al., 2008; Saykin et al., 1991), excepting smooth pursuit eye movements (Park & Holzman, 1993). This is further supported by our finding that although SZ decrease their rate of spatial errors faster than the other two groups, all three groups finish the repetitions of a sequence with roughly the same number of errors. Park et al. (2008)
found significant differences between CO and SZ for online imitation in a similar task to the one described here, but they had trained experimenters score the imitations, as opposed to using our methodology.

In addition, there are two distinct reasons for the lack of difference in spatial errors between the groups, as well as low general number of errors in the present study. First, the task in the current study involved only one or two gestures per sequence. There are fewer errors in this study compared to Studies 1 and 2, both of which involve longer sequences, and it can therefore be inferred that longer sequences would result in the production of more spatial errors, which might then discretely separate the groups. Also, it has been shown through a series of studies that SZ are not impaired in attention compared to CO (J. M. Gold et al., 2006). This lack of impairment in attention, coupled with the lack of motor impairment (Saykin et al., 1991; Foerde et al., 2008; Pedersen et al., 2008), explains the lack of differences between the groups in the current study.

Differences between the subject groups in the current study come from the temporal aspect of imitation. Both the SZ and BP groups are slower than the CO group when starting their imitation from memory, with the SZ group taking the longest during the first repetition. The temporal differences between SZ and CO, especially before beginning the imitation, suggests an impairment in memory encoding, in line with previous studies (J. M. Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997; Fuller, Luck, McMahon, & Gold, 2005).

To expand the current study, it would be beneficial to use stimulus sequences with more gestures, as introducing more gestures introduces more opportunities for subjects to make spatial errors and see temporal discrepancies. Elvevag et al. (2003) argues that there is a limited span for serial order in SZ, which is due to impaired memory encoding rather than a specific serial order memory problem. Though we find some serial order errors, there are simply not enough gestures in the sequences to elicit sequence errors (Elvevag, Egan,
& Goldberg, 2000), gesture omissions on the order seen in verbal recall tasks (Elvevag, Weinberger, & Goldberg, 2001) or serial position deficits as found by Elvevag et al. (2003). Finally, in an effort to reconcile our findings with Park et al. (2008), a current study is being conducted in our lab in which the ratings of human observers will be compared to the findings of our algorithm in the hopes that we can make our algorithm accurately reflect the natural biases of human observers.
General 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 & 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 allows us to compare an imitation of any multi-dimensional 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.

The first study demonstrated 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 visuo-spatial memory.

In addition, we illustrated that our algorithm could identify chronometric properties of imitation, with longer pre-movement latencies observed during subjects’ early imitations. This presumably reflects a learning effect in that the effort, be it cognitive or motor, needed by the subject decreases with repetition, as the subject learns the sequence. Not only do experts take less time overall in their reproductions compared to nonexperts, but patients take longer than controls before beginning their imitation, with all groups decreasing their pre-motor latencies the more they imitate a specific stimulus. Further, Study 2 confirms the results of Study 1 that the more complex a stimulus is, the more spatial errors are made in reproduction, and the longer it takes subjects to reproduce the stimulus.

Spatial errors, most notably gesture errors, followed the same pattern as the pre-motor latencies. That is, across all three studies, sequence types and conditions that produced shorter times (pre-movement latencies, hold times and transition times) also had fewer spatial errors compared to the other sequence type(s), as shown in Figures 6A and B, 8, 9, 10, and 11. In all studies, the number of errors for each sequence type converge after just a few repetitions, eliminating differences between the sequence types. These findings suggest
that subjects have learned the stimulus more successfully after viewing and imitating the
stimulus more than once compared to viewing and imitating it only once, and thus the
subjects are better consolidating the stimulus in memory the more they see it.

More experiments should be performed to test this theory of consolidation. One
possibility is that after viewing a stimulus multiple times, subjects develop a strategy in
which they label each component of the stimulus. This is supported by the initial shorter
hold times by the ASL-experienced subjects on only those stimuli involving gestures with
which they were familiar (the signed-sequences), but not on the unfamiliar gestures. One
idea would be to use manual or articulatory suppression, or a combination, to study the
temporal changes in an expert group. This would determine if these impediments would
alter performance to experts’ level on the stimuli with which they have no expertise by
affecting an ASL rehearsal loop (Wilson & Emmorey, 1997), similar to the Baddeley and
Hitch (1974) phonological loop. Similarly, developing a quick encoding strategy for the
stimuli could result in faster consolidation and learning, such that schizophrenic patients
achieve the level of normal controls from the beginning (and nonexperts reach the level of
experts), as opposed to after multiple repetitions.

The decrease of spatial errors and improvement in temporal performance with repe-
tition can be attributed to stimulus priming (Goldstone, 1998). In that article, Goldstone
claims that ”people are better able to perceptually identify unclear or quickly presented
stimuli when they have been previously exposed to them” (p 592). As such, the faster
times of the ASL-experienced group on the familiar gestures, that is, in the signed-sequences
in Study 2, would be due to prior stimulus imprinting, whereas they would not have the
same priming for the nonsense-sequences. Likewise, the ASL-naive group would have no
prior experience with either stimulus type, and thus any priming would occur during the
initial stimulus presentations, and again, we saw decreases with repetition corresponding
to stimulus complexity, the number of changes in digit flexion and extension throughout
the stimulus. This also holds true for the subject groups in Study 3, such that subjects
improved on multiple analyses with repetition, even though the patient groups may initially
be slower or have more errors than the control group. Differentiation, another of the Gold-
stone mechanisms of perceptual learning (Goldstone, 1998) is also being utilized, at least in
Study 2. That is, those who have prior ASL experience are more easily able to discriminate
individual components of a stimulus model for the purposes of reproduction, compared to
those who lack knowledge of ASL.

One way in which we can look at the differences between expertise to determine the
types of learning taking place is to look at cortical activations for experts versus nonexperts.
Newman, Bavelier, Corina, Jezzard, and Neville (2002) showed that in native bilingual sign-
ers there is activity in the right hemisphere angular gyrus that is not found in hearing signers
who acquired ASL after puberty. Given that the angular gyrus is involved in language and
cognitive processes, it is therefore likely that this difference in activity would result in better
performance in Study 2 if we were to test deaf, native signers. Yet Keehner and Gathercole
(2007) argue that cognitive adaptations can occur with any sign language experience, no
matter when acquired. The results of Study 2 show that there is some advantage afforded
experts, provided the stimuli to be learned are directly related to the field of expertise.

In addition, it has been argued that rather than having discrete finger somatotopy in
the primary motor cortex (M1), there is instead an overlapping digit representation in M1
(see Sanes & Schieber, 2001 for a summary, and Hlustík, Solodkin, Gullapalli, Noll, & Small, 2001; Rao et al., 1995; Sanes, Donoghue, Thangaraj, Edelman, & Warach, 1995 for studies on the overlap). It is therefore possible that the representations are different for expert signers than for nonsigners. Schieber and Santello (2004) state that "neural control allows adjacent fingers to move together when they are not acting on keys but actively dissociates them when one must act without the other", but it is possible that when adjacent fingers are formed into a known sign, the activation differs compared to when they are formed into a meaningless sign, as would fit Kleinschmidt, Nitschke, and Frahm (1997)’s theory of qualitative predominance of certain hand configurations over others in M1. They found a mediolateral progression of the digits in M1 by comparing individual digit movements. As Kleinschmidt et al. (1997) did not use more complex movements as found in sign language, it would be advantageous to see if there is a difference in activation in M1 for both signs and nonsigns, and even for each dez (handshape configuration) in sign language (Stokoe, 1960). Additionally, we can look at the difference in activation between experts and nonexperts. It may be that there are differences in the inferior frontal gyrus, inferior parietal lobule, and temporo-parieto-occipital junction between experts and nonexperts, as have been found when comparing subjects with intact imitation ability to those with impaired imitation ability (Goldenberg & Karnath, 2006).

It may also be the case that biomechanical constraints of the hand influence errors that made in imitation. 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 are more in motion than the non-adjacent digits, especially when the intended finger is a highly coupled finger (Fish & Soechting, 1992; Hager-Ross & Schieber, 2000). As such, our algorithm could be improved by studying the relationship of gesture-level errors in imitation to the biomechanical constraints of the hand. It is likely that biomechanical coupling of the hand results in more flexion errors between highly coupled, adjacent digits than more individuated digits (Loehr & Palmer, 2006; Schieber & Santello, 2004).

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 the right hand, and characterizes the imitation’s various chronometric properties. In each study presented here, a pre-study 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 adequate for our current experimental paradigm (where digits in the model were either fully flexed or extended), this approach would not be appropriate 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.

In addition, though we can quantify many aspects of imitation, the interpretations of some measures are not as unambiguous as one would like. For example, longer premotor latencies could reflect an increased cognitive load, a problem with memory consolidation,
problems with motor planning, slowed motor skills, or a combination of the above. Similarly, in the reproduction of some sequences, subjects transitioned from one gesture to the next more quickly than when reproducing other sequences. Yet it is unclear as to whether this means that subjects are concerned about forgetting the gestures in a sequence and thus try to complete the reproduction as quickly as possible, or if, conversely, subjects are more sure of those sequences, or, perhaps that subjects settle their fingers into position more quickly in those sequences than others. We cannot fully interpret these transition times until we decouple stimulus complexity from expertise. Thus, while our algorithm can quantify these spatial and temporal measures, it is imperative that further studies be designed to understand what their values mean, so as to improve learning and imitative ability. For the same reason, future studies must also be designed to separate motor effort from cognitive performance.

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 studies 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, in addition to 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.

The studies presented here represent a first step toward a quantitative analysis of complex human imitation. The methodology used in the three studies has broad implications for cognitive neuroscience and neuropsychology. As shown, the methodology can be used to study expertise and impairment in reproduction and other forms of imitation. With simple extensions the methodology can study more complex and intricate movements, such as breaking down the swing of each member of a baseball team. From the information about imitation gleaned from our methodology, we are hopeful that we can discover more efficient methods of teaching and learning.
References


