Evaluation Metrics and Results of Human Arm Movement Imitation

Marc Pomplun[1] and Maja Matarić[2]

[1] University of Toronto, Ontario, Canada M5S 3G3,
marc@psych.utoronto.ca <mailto:marc@psych.utoronto.ca>
[2] University of Southern California, Los Angeles, CA 90089-0781, USA,
mataric@cs.usc.edu <mailto:mataric@cs.usc.edu>
WWW home page: <http://www-robotics.usc.edu/~maja>

Abstract. We present a psychophysical study of human arm movement imitation, and an approach to analyzing the resulting data, which is general enough to be applied to human or humanoid movement analysis. We describe a joint-space based segmentation and comparison algorithm that allows us to evaluate the performance of 11 different subjects performing a series of arm movement imitation tasks. The results provide analytical evidence for the strong interference effects of simultaneous rehearsal during observation. Additionally, the results also demonstrate that repeated imitation in these tasks did not affect the subjects' performance.

1 Introduction

Our interest is in gaining insight into the mechanisms behind imitation, the ability to repeat an observed behavior as well as to learn arbitrary new skills, i.e., skills that are irrelevant to the goal, by observation [2]. So called "true imitation" is thought to be a complex mechanism, since it is found in very few species, while mimicry is more common [17]. We have proposed a model of imitation [14, 9] based on evolutionarily older substrates including motor primitives as the basis for motor control and the mirror system [8, 19, 20] as the basis for a direct sensory-motor mapping for mimicry. Our other work [14, 9, 4, 23, 1] is aimed at developing and implementing this model on a variety of synthetic humanoid platforms, in order to validate the model on real-world tasks.

In this paper, we focus on a set of psychophysical imitation experiments we performed in order to gain further insight and constraints for our imitation model. We gathered arm movement data from a collection of normal subjects imitating video demonstrations, and have developed a set of analysis tools for evaluating the quality of the resulting imitation. Evaluation of imitation is an open problem, and our evaluation metrics are meant as a general tool for both human and robot imitation and general motor control evaluation.

The rest of this paper is organized as follows. Section 2 summarizes previous research and develops the questions investigated by the present study. The resulting experimental design is described in Section 3. In Section 4, the components of an appropriate metric for measuring distances between pairs of trajectories are specified, and in Section 5 the contribution of each component to the performance of the metric is analyzed. Section 6 provides the experimental results, including a statistical analysis of the factors influencing the subjects' imitation performance. In Section 7, these results are discussed, and the paper is summarized in Section 8.

2 Motivation

Imitation is a powerful tool of cultural skill transfer and interaction [2]. Thus, it is potentially of great value to robotics, where control of and interaction with highly-complex humanoid robots that are currently becoming available (Honda P3, NASA Robonaut, Sarcos full-body, etc.) are open problems.

Our earlier work [12, 15] in the area of psychophysics of imitation involved studying the attentional behavior of people watching movements for subsequent imitation. Specifically, we recorded the visual fixations of 40 subjects (half male, half female), wearing an eye-tracker and watching videos of finger, hand, and arm movements. We constructed the experiments so as to address the following questions:

1. Is there a difference between watching a movement with the intention to imitate and just watching?
2. When watching with the intention to imitate, what features are fixated on?

Our data demonstrated conclusively that the answer to the first question was negative, and the answer to the second was revealing. Subjects fixated at the end-point (finger or hand, or if a pointer was present, the tip of the pointer), regardless of whether they were just watching or intending to imitate. The only difference we observed was pupil dilation, which was larger in the imitation condition, presumably indicating cognitive processing.

These results were important because they provided insight into human attention in imitation, and an indication that people most likely use internal models of behavior to aid movement understanding and imitation, rather than relying on detailed observation [15].

In spite of their fixation at the end-point, subjects in the first set of experiments were able to successfully imitate most of the demonstrated movements. In the experiments presented here, we aimed to answer more precise questions about the quality of the subjects' imitation under various conditions, by recording the demonstrator's and the subjects' movements for subsequent analysis and comparison.

This presented an important evaluation challenge. What is the most effective means of evaluating arm movements across subjects and tasks? The evaluation tools we developed were based on trying to answer this question, in order to be able to make quantitative statements about human movement, resulting from imitation or otherwise. We are not aware of any other work aimed directly at analytically evaluating imitated behavior, and the work presented here is at its very beginning.

Research into copying written 2D characters, drawings, and trajectories has been done [3, 10]. In our previous work on evaluation of motor behavior (but not imitation behavior) [16], we explored end-effector jerk as a quantitative measure of motor behavior, but found it to be overly sensitive to specific parameters of each implementation. Other measures, such as minimum torque change [22] and minimum energy [18], have also been explored. None of these alone provides a sufficient description of the quality of a movement, and in particular of a motor imitation, where the underlying intentions of the demonstrator, however complex, may be relevant. Depending on the nature of the task, imitation can be as simple as achieving the observed task (so-called "task level" imitation [21]), something that is quite easily evaluated, to being as complex as conveying the underlying message of a dance or a gesture. The latter has so far only been studied qualitatively.

Studies evaluating task-irrelevant movements [5–7] compared human movements to the output of computer vision systems tracking and modeling these movements by using a simple mean square error metric. They demonstrated that in many cases even pairs of movements with large MSE were still perceived as identical by human observers. This motivates the use of more sophisticated metrics to quantitatively assess the similarity of limb trajectories.

With the growing interest and focus on humanoid robotics, the need for control and thus evaluation of behavior of such systems in particular for the purposes of interaction with humans is of key importance. It is toward that end that we conducted the psychophysical study and then turned toward the analysis of its data.

In the next section we describe the experimental design of the study we conducted.

3 Experimental Design

Our experiments were conducted at the National Institutes of Health Resource for the Study of Neural Models of Behavior, at the University of Rochester.

3.1 Subjects

The subject pool consisted of 11 right-handed subjects, 7 female and 4 male. One of the authors (M. M.), who is also right-handed, participated in a separate session to produce reference data for comparison, i.e., the correct demonstrated movements.

3.2 Stimuli

The stimuli consisted of 20 video clips, each showing an individual sequence of human arm movements. These movements had been performed by one of the authors (M. M.) and video recorded. The video clips presented her arm moving in front of a black background; each clip was about 3 to 5 seconds long. All sequences started with a straightened arm and a clenched fist and typically involved movements of the shoulder, elbow, wrist, and fingers. The stimuli involved no external objects or recognizable patterns, so they can be thought of as goal-independent; the only goal the subjects had was to repeat, i.e., imitate, the presented behavior as accurately as possible.

3.3 Apparatus

Stimuli were presented on a 22-inch monitor at a viewing distance of 100 cms. Arm movements were tracked with the FastTrak motion tracking system. For this purpose, four position sensors were fastened to the subject's right arm, one in each of the following four positions (see Figure 1):

1. at the center of the upper arm,
2. at the center of the lower arm,
3. immediately above the wrist, and
4. on the upper phalanx of the middle finger.

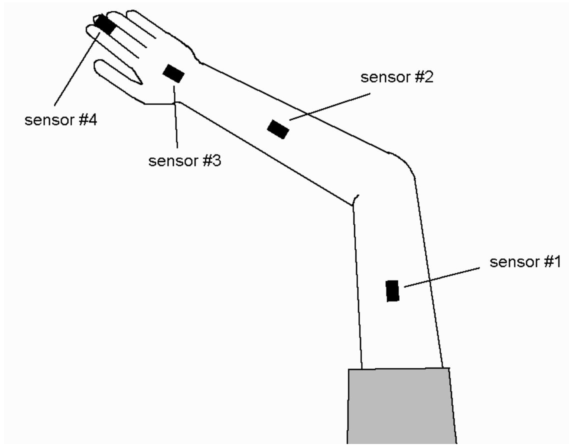

Fig. 1. Arrangement of the four FastTrak sensors on the subject's right arm

Each sensor was about $2\times1\times1$ cm in size. We used elastic bands to hold the sensors and their cables in place. The cables were taken to the subject's right shoulder and then connected to the FastTrak processing computer. The spatial position of each sensor was measured every 34 ms with an accuracy of about 2 mm in each dimension.

3.4 Procedure

Prior to each recording trial, one of the 20 stimuli was shown to the subject. We divided the trials into two conditions: "rehearsal" and "no rehearsal". In the "rehearsal" condition, the subject imitated the sequence of movements while they were watching it on the screen, i.e., performed simultaneous rehearsal. In the "no rehearsal" condition, the subject's right arm rested on a desk while the video clip was playing.

To initiate the imitation process after each stimulus presentation, the experimenter gave the verbal signal "ready". The subject then straightened out his or her arm horizontally towards the monitor screen and clenched the fist. Next, the subject received the instruction "imitate" and started to imitate the previously viewed sequence of movements as precisely as possible. When finished imitating, the subject said "done" and put the arm back into the resting position. Between the "imitate" and "done" signals, the subject's arm movements were recorded by the FastTrak system.

In the trials reported here, each subject was presented with four different stimuli, each of which was shown and imitated three times in succession. Two of the four stimuli were imitated in the "rehearsal" condition, the other two were imitated in the "no rehearsal" condition. The sequence of stimuli was counterbalanced across subjects.

To obtain reference data for correctly imitated movements, M.M. participated in a separate session in which she imitated the 20 sequences of movements that the 11 subjects had previously imitated. Each of the corresponding 20 stimuli was shown to her three times, each time followed by her imitation, resulting in a total of 60 reference trajectories being recorded.

4 Analysis Methods

The goal of our analysis was to evaluate any given imitation. Toward this end, we developed a metric for the distance (or difference) between arm movement trajectories. Instead of comparing the spatial Cartesian coordinates of the four sensors, we chose to apply evaluation in joint-space, because joint angles are independent from the length of the subject's limbs, thus compensating for individual physical differences. We derived the following five angles from the Cartesian coordinates of the four position sensors:

1. horizontal shoulder angle
2. vertical shoulder angle
3. shoulder rotation (along the axis going from the shoulder to the elbow)
4. elbow
5. wrist (including finger angles)

Starting from a simple mean square error (MSE) metric between corresponding joint angle samples, we developed a more sophisticated and adequate metric by successively adding new components that reduced noise-sensitivity. As a result, such a distance metric cannot be sensitive to all of the many diverse aspects of imitation performance. It is, however, a straightforward, quantitative approach meeting the demands of this first study to evaluate imitated behavior.

In the present data, the sampling rate was a constant 34ms. Consequently, the following explanation of the metrics will refer to sample indices instead of time-codes. This in no way restricts the generality of the described methods; trajectories with differing or shifting sampling rates can still be analyzed after conducting a time scaling operation we describe below.

To establish notation, let us assume that we want to compare two arm movement sequences (trajectories) α and β containing $(T_\alpha + 1)$ and $(T_\beta + 1)$ samples respectively for each of the J joint angles. We will refer to the samples as $\alpha_t^{(j)},\ j = 1, \ldots, J,\ t = 0, \ldots, T_\alpha$ and $\beta_t^{(j)},\ j = 1, \ldots, J,\ t = 0, \ldots, T_\beta$. Our aim is to define a distance metric $d(\alpha, \beta)$ yielding lower values for more similar trajectories α and β.

We now describe four comparison methods of increasing effectiveness: raw trajectory comparison, time scaling, separate segmentation, and combined segmentation.

4.1 Raw Trajectory Comparison

The mean square distance between corresponding angles in α and β trajectories, across joints and samples, is a simple and straightforward difference metric that ignores possible additional samples in the longer trajectory:

$$d(\alpha, \beta) = \sum_{t=0}^{\min(T_\alpha, T_\beta)} \sum_{j=1}^{J} (\alpha_t^{(j)} - \beta_t^{(j)})^2 \tag{1}$$

In order to compensate for differences between α and β trajectories in the starting position of the arm, before calculating (1), time-invariant offsets are added to all joint angles in both trajectories so that all angles start at 0.

The most evident shortcoming of this metric is that a slight difference in velocity can cumulatively grow into a large distance between the two trajectories being compared. To eliminate this undesirable feature, a time scaling procedure can be performed prior to the calculation of mean square distances.

4.2 Time Scaling

In order to derive a metric that is invariant to the absolute speed of movement and accounts for all data in both α and β trajectories, the shorter trajectory, say α, is expanded to contain as many samples as its counterpart. To do this, a continuous function $\tilde{\alpha}^{(j)}(\tilde{t})$ interpolating α is needed:

$$\tilde{\alpha}^{(j)}(\tilde{t}) = \alpha_t^{(j)} \ \forall \ j = 1, \ldots, J,\ \ t = 0, \ldots, T_\alpha,\ \ \tilde{t} = t \tag{2}$$

For most kinds of trajectories, spline interpolation is an adequate method. Based on the interpolation, the samples in α can be rearranged to match the number of their counterparts in β:

$$\alpha_{t,new}^{(j)} = \tilde{\alpha}^{(j)}(\frac{t \cdot T_\alpha}{T_\beta}) \ \forall \ j = 1, \ldots, J,\ \ t = 0, \ldots, T_\beta \tag{3}$$

$$T_{\alpha,new} = T_\beta \tag{4}$$

After time scaling, the metric $d(\alpha, \beta)$ can still be calculated as the mean square distance between α and β, according to Equation (1) above.

4.3 Separate Segmentation

Since in the present context trajectories are sequences of movements, the metric should benefit from identifying the elementary trajectory segments and comparing them individually. A useful indicator of transitions between successive segments in a trajectory α is the velocity $\dot{\alpha}_t$ calculated across all joint angles:

$$\dot{\alpha}_t = \sqrt{\sum_{j=1}^{J} (\alpha_t^{(j)} - \alpha_{t-1}^{(j)})^2}, \quad t = 1, \ldots, T_\alpha \tag{5}$$

When examining velocity histograms within any of the recorded trajectories, we always find the highest and widest peak at a very low velocity. This peak reflects the almost motionless phases in the trajectory that occur between successive segments and also at the beginning and the end of each trajectory. The basic idea of our segmentation algorithm is to determine contiguous phases in the trajectory with velocities above a certain threshold $\theta_{\dot{\alpha}}$. Such contiguous phases exceeding a specific minimum duration are considered to be the segments of the trajectory.

We determined empirically that an appropriate value for the velocity threshold can be estimated from the position $\dot{\alpha}_{max}$ of the peak in the velocity histogram:

$$\theta_{\dot{\alpha}} = 2.5 \cdot \dot{\alpha}_{max} \tag{6}$$

For the present trajectories, the minimum duration for segments was set to 200 ms.

In many cases, the segmentation of the α and β trajectories to be compared leads to different numbers of segments. To calculate the distance $d(\alpha, \beta)$ in such cases, the segmentation of the trajectory with more segments, say α, is adjusted, i.e., the number of segments in α is reduced to equal those in β through the use of an iterative algorithm. In each step, the algorithm combines two successive segments in α into one segment. The segments to be combined are those whose combination leads to the smallest sum of temporal deviations between corresponding segment boundaries in α and β. This procedure is repeated until the same number of segments in α and β is reached.

Within corresponding segments, for example the first segments in α and β, time scaling is carried out as shown in Equations (2) to (4). Subsequently, the metric $d(\alpha, \beta)$ is calculated as the mean square distance across all samples and segments, which corresponds to Equation (1). We call this the "separate segmentation" algorithm, because the segmentations for the trajectories α and β to be compared are carried out separately before their distance $d(\alpha, \beta)$ is determined.

4.4 Combined Segmentation

Although the threshold estimation used in the "separate segmentation" algorithm above works well for most trajectories, there are cases in which two rather similar trajectories are segmented clearly differently, which leads to an inappropriately large distance between them. To avoid such distortions in the metric, a "combined segmentation" algorithm is employed, which finds the two thresholds that lead to the best match between the two trajectories being compared.

Instead of estimating an appropriate threshold $\theta_{\dot{\alpha}}$ for trajectory α, the "combined segmentation" algorithm determines a maximum threshold $\theta_{\dot{\alpha},max}$ limiting the range of thresholds to be examined. Empirical investigations found the following equation to yield best results:

$$\theta_{\dot{\alpha},max} = 3.5 \cdot \dot{\alpha}_{max} \tag{7}$$

After determining $\theta_{\dot{\alpha},max}$ and, analogously, $\theta_{\dot{\beta},max}$ for the trajectories α and β, the algorithm systematically performs segmentations with different combinations of velocity thresholds $\theta_{\dot{\alpha}}$, ranging from 0 to $\theta_{\dot{\alpha},max}$, and $\theta_{\dot{\beta}}$, ranging from 0 to $\theta_{\dot{\beta},max}$. For the trajectories in our data set, it was sufficient to use 20 levels for each of the two variables, resulting in 400 combinations to be tested. For each combination of segmentations, the mean square distance between α and β was calculated using the "separate segmentation" algorithm. $d(\alpha, \beta)$ was determined as the minimum distance found within these 400 combinations, i.e., the distance for the best-matching segmentations of α and β.

5 Method Validation

In order to evaluate the performance of the distance metric and the contribution of each of its four components, we compared the *intra*-stimulus distances to the *inter*-stimulus distances in our set of 60 reference movements performed by M.M. *Intra*-stimulus distances are distances between pairs of trajectories for the same stimulus, i.e., imitations of the same movement. Since we had three imitations for each stimulus, there were three trajectories and thus three distances between trajectory pairs (1-2, 1-3, 2-3), leading to a total of 60 intra-stimulus distances. The term *inter*-stimulus distances refers to distances between pairs of trajectories for different stimuli, i.e., imitations of different movements. Inter-stimulus distances were measured between each stimulus S and those 57 trajectories corresponding to stimuli other than S, resulting in a total of 1710 inter-stimulus distances. We would expect an accurate and reliable metric to detect larger differences between mean intra-stimulus distances than between inter-stimulus distances. Similarly, we would expect the metric to detect smaller standard deviations in the mean intra-stimulus distances than in the inter-stimulus ones.

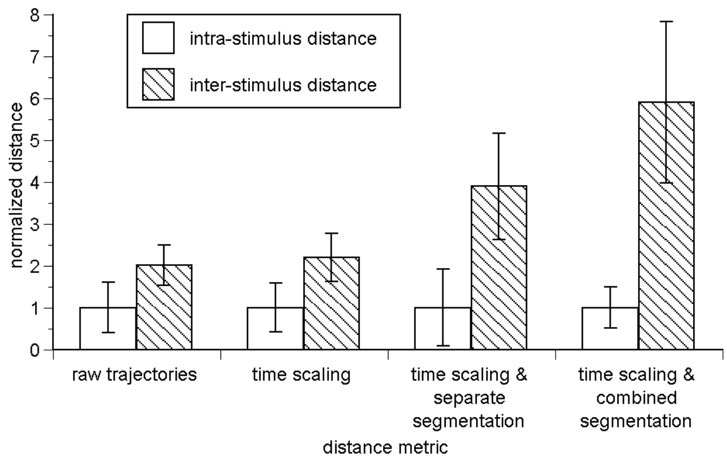

Fig. 2. Mean intra- and inter-stimulus distances and their standard deviations as produced by the four variants of the metric

Figure 2 demonstrates the performance of the metric by showing intra- and inter-stimulus distances for each of the four variants of the metric (raw samples, time scaling, time scaling and separate segmentation, time scaling and combined segmentation). For each variant, these data were normalized in such a way that the mean intra-stimulus distance was always 1. As can clearly be seen, adding a component always led to an improvement in the performance of the metric, with the last component, combined segmentation, having the most pronounced effect.

The four stages of the distance metric were also evaluated in a classification test. For the first trajectory α of the first stimulus, for example, we calculated its mean distance from the second and third trajectory for each stimulus, resulting in 20 distance values. Trajectory α was then assigned to the stimulus whose second and third trajectories had the shortest mean distance from α. In our example, if the trajectory was assigned to the first stimulus, a correct classification was made. This classification test was performed on all 60 stimuli. Second trajectories were compared to first and third trajectories, and third trajectories were compared to first and second trajectories.

Figure 3 presents the number of incorrect classifications produced by each of the four variants of the metric. As can be seen, each additional component reduces the number of incorrect classifications, with the most sophisticated variant, time scaling and combined segmentation, performing perfect classification on the given set of trajectories. Taken together, the above two evaluations of the metric reveal both the usefulness of all its components and the remarkably good performance of the final combined metric.

Next, we discuss the results of applying this metric to the human imitation data.

6 Imitation Results

Having developed and validated the comparison and evaluation metric for the arm movement trajectories, we applied it to the subjects' data. The mean distance between the subjects' trajectories and the three corresponding

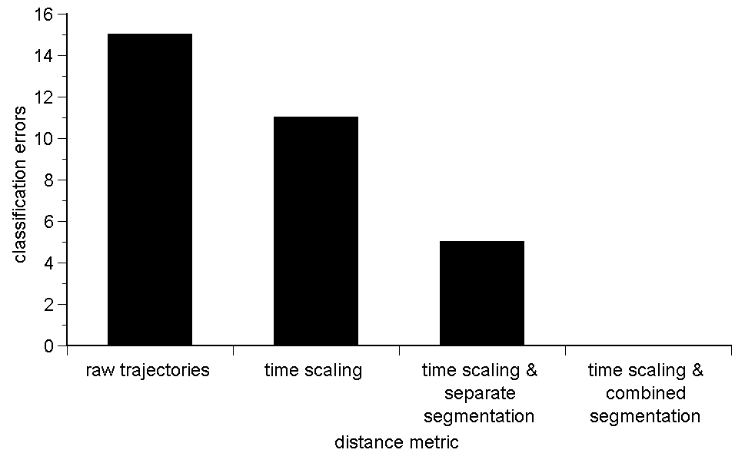

Fig. 3. Incorrect classifications (out of 60) made by the four variants of the metric

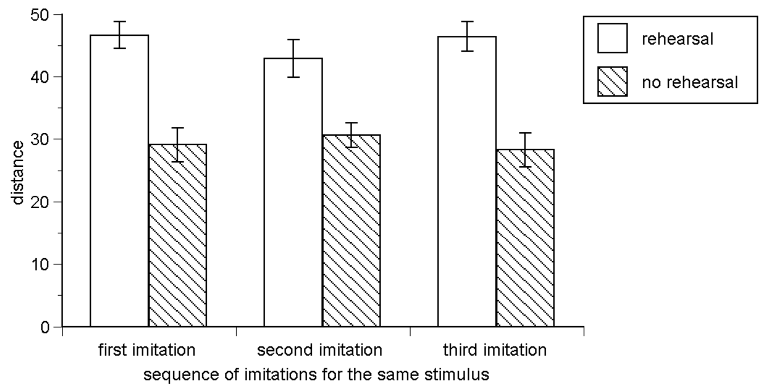

Fig. 4. Distance between subjects' and reference trajectories across experimental conditions

reference trajectories for each imitation were analyzed separately for six different conditions: (rehearsal vs. no rehearsal) $\times$ (first, second, and third imitation of the same stimulus).

Figure 4 plots the results of this analysis. A 2 (rehearsal vs. no rehearsal) $\times$ 3 (sequence: first, second, and third imitation) analysis of variance (ANOVA) revealed a significant effect of rehearsal on the subjects' imitation performance, $F(1, 10) = 270.59; p < .001$. When subjects rehearsed the sequence of movements while viewing the stimulus, their performance in the subsequent imitation was substantially *weaker*, i.e., the distance between their trajectories and the reference trajectories was greater than without rehearsal. The mean with-rehearsal error was 45.30, while the mean no-rehearsal error was 29.49.

In contrast, the subjects' performance was not influenced by the imitation sequence, $F(1, 10) < 1$. There was no significant difference in performance between the first, second, and third imitation of the same stimulus. The associated mean errors were 37.83 for the first, 36.73 for the second, and 37.62 for the third imitation of the same stimulus.

No significant interaction between the two factors was found.

7 Implications and Applications

Our results demonstrate conclusively that imitation with simultaneous rehearsal leads to worse performance than imitation without such rehearsal. Specifically, simultaneous rehearsal interferes detrimentally with retention necessary for subsequent imitation, regardless of the delay between the presentation and imitation. This result is consistent with brain imaging studies demonstrating interference in other working memory tasks [11]. Furthermore, this analytically confirms our own earlier qualitative finding that demonstrated the same effect; our subjects in the earlier experiment [15] were given the option to rehearse, but all chose not to after one or two attempts, reporting a subsequent inability to remember the observed stimulus. Perhaps more importantly, the interference data indicate a more general effect: the detrimental interference of simultaneous rehearsal is significant regardless of the delay between the presentation and imitation.

Our results also demonstrate that repeated imitation does not improve the subjects' performance in this particular task. We postulate that the nature of the arm movements used as stimuli was simple enough that feedback from physical execution of the movement was not helpful. Had the same stimuli been presented repeatedly between imitation performances, i.e., had subjects been given the opportunity to learn through repeated trials, we expect that at least some improvement in performance would have been observed. This result would not be surprising or novel, however.

Given the rather straightforward nature of the movements used in the experiment, we can conclude that subjects generated the motor imitation likely by resorting to generic movement patterns that were familiar, and did not improve those through repeated execution. This is consistent with our finding that even when the subjects were made to fixate away from the end-point, as they would naturally do [15], but instead look at the elbow, the quality of their subsequent imitation was *not* drastically affected. Additionally, this fits well within our model of imitation [14], which postulates that all observed human movements are mapped directly (through the mirror system) onto a combination of known motor programs or primitives, and it is the combination of those primitives, through rehearsal, that generates novel motor skills.

The evaluation approach we describe is general and can be applied to any joint-space representation of movements. We are currently using it to compare human performance to humanoid avatar performance in the context of implementing and evaluating an imitation model for complex humanoid robots. As discussed earlier, this is just one means of quantitatively evaluating and comparing joint-space trajectories. It does not address the more complex issue of task-level imitation. However, it does provide a reliable and robust quantitative metric for movement evaluation, for imitation and otherwise.

8 Summary

In this paper we presented a psychophysical study of human arm movement imitation, and an approach to analyzing the resulting data. We described a joint-space based segmentation and comparison algorithm that allowed us to evaluate the performance of 11 different subjects on a series of arm movement imitation tasks. The results provided analytical evidence for the strong interference effects of simultaneous rehearsal during observation. Additionally, the results also demonstrated that repeated imitation in these tasks did not affect the subjects' performance.

Acknowledgements

The research described here was supported by the NSF Career Grant IRI-9624237 to M. Matarić. The data used for this analysis were gathered by M. Matarić and M. Pomplun in a joint interdisciplinary project conducted at the National Institutes of Health Resource for the Study of Neural Models of Behavior, at the University of Rochester. The authors are grateful to Prof. Dana Ballard and Prof. Mary Hayhoe for hosting their experiments and to Prof. Jeff Pelz for enduring help during the experiments as well as afterwards. The authors thank Ajo Fod for his help in finalizing the submission.

References

1. Billard, A. & Matarić, M.J.: Betty: Robot, play with me! Robot: O.K. How do we play? Betty: You watch me and do like I do. Look!, *in*. 'Proceedings, Workshop on Interactive Robotics and Entertainment (WIRE-2000)', Pittsburgh, 2000.
2. Byrne, R. W. & Russon, A. E.: Learning by Imitation: a Hierarchical Approach, *The Journal of Behavioral and Brain Sciences* **16**, 3, 1998.
3. Cooke, S., Kitts, B., Matarić, M.J. & Sekuler, B.: Delayed and Real-Time Imitation of Complex Visual Gestures, *in* 'Proceedings, International Conference on Vision, Recognition, Action: Neural Models of Mind and Machine', Boston University, 1997.
4. Fod, A., Matarić, M.J. & Jenkins, O.C.: Automated Extraction of Primitives for Movement Classification, *in* 'Proceedings, First IEEE-RAS International Conference on Humanoid Robotics (Humanoids-2000)', MIT, Cambridge, MA, Sep 7-8, 2000.

5. Goncalves, L., Di Bernardo, E., Ursella, E. & Perona, P.: Monocular tracking of the human arm in 3D, *in* ‘Proceedings, ICCV 95’, 1995.
6. Goncalves, L., Di Bernardo, E. & Perona, P.: Reach Out and Touch Space (Motion Learning) *in* ’Proceedings FG98, International Conf. on Face and Gesture Recognition’, 1998.
7. Goncalves, L., Di Bernardo, E. & Perona, P.: Monocular perception of biological motion in Johansson Displays *in* ’Proceedings of ICCV 99, Corfu, Greece’, 1999.
8. Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C. & Rizzolatti, G.: ‘Cortical Mechanisms of Human Imitation’, *Science* **286**, (1999) 2526-2528.
9. Jenkins, O.C., Matarić, M.J. & Weber, S.: Primitive-Based Movement Classification for Humanoid Imitation, *in* ‘Proceedings, First IEEE-RAS International Conference on Humanoid Robotics (Humanoids-2000)’, MIT, Cambridge, MA, Sep 7-8, 2000. Also IRIS Technical Report IRIS-00-385, 2000.
10. Kitts, B., Cooke, S., Matarić, M.J. & Sekuler, R.: Improved Pattern Recognition by Combining Invariance Methods, *in* ‘Proceedings, International Conference on Vision, Recognition, Action: Neural Models of Mind and Machine’, Boston University, 1997.
11. Klingberg, T.: ‘Concurrent Performance of Two Working Memory Tasks: Potential Mechanisms of Interference’, *Cerebral Cortex* **8**(2), (1998) 593-601.
12. Matarić, M.J. & Pomplun, M.: What do People Look at When Watching Human Movement?, Technical Report CS-97-194, Brandeis University, 1997.
13. Matarić, M. J.: Learning Motor Skills by Imitation, *in* ‘Proceedings, AAAI Spring Symposium Toward Physical Interaction and Manipulation’, Stanford University, 1994.
14. Matarić, M.J.: Sensory-Motor Primitives as a Basis for Imitation: Linking Perception to Action and Biology to Robotics, *in* C. Nehaniv & K. Dautenhahn, eds, ‘Imitation in Animals and Artifacts’, The MIT Press, 2000.
15. Matarić, M. J. & Pomplun, M.: ‘Fixation Behavior in Observation and Imitation of Human Movement’, *Cognitive Brain Research* **7**(2), (1998) 191-202.
16. Matarić, M. J., Zordan, V. B. & Mason, Z.: Movement Control Methods for Complex, Dynamically Simulated Agents: Adonis Dances the Macarena, *in* ‘Autonomous Agents’, ACM Press, Minneapolis, St. Paul, MI, (1998) 317-324.
17. Moore, B.R: The Evolution of Imitative Learning, *in* C. M. Heyes & B. G. Galef, eds, ‘Social Learning in Animals: The Roots of Culture’, Academic Press, New York (1996) 245-265.
18. Nelson, W.L.: ‘Physical Principles for Economies of Skilled Movements’, *Biological Cybernetics* **46**, (1983) 135-147.
19. Rizzolatti, G. & Arbib, M.: ‘Language within our grasp’, *Trends in Neuroscience*, 1998.
20. Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Perani, D. & Fazio, F.: ‘Localization of Grasph Representations in Humans by Positron Emission Tomography: 1. Observation Versus Execution’, *Experimental Brain Research* **111**, (1996) 246-252.
21. Tomasello, M., Kruger, A. C., & Rather, H. H.: ‘Cultural Learning’, *The Journal of Behavioral and Brain Sciences* **16**(3), (1993) 495-552.
22. Uno, Y., Kawato, M. & Suzuki, R.: ‘Formation and Control of Optimal Trajectory in Human Arm Movement-Minimum Torque-Change Model’, *Biological Cybernetics* **61**, (1989) 89-101.
23. Weber, S., Jenkins, O. C. & Matarić, M.J.: Imitation Using Perceptual and Motor Primitives, *in* ‘Proceedings, The Fourth International Conference on Autonomous Agents (Agents 2000)’, Barcelona, Spain, June 3-7, 2000, 136-137.