Control and Imitation in Humanoids

Maja J. Matarić, Odest C. Jenkins, Ajo Fod, and Victor Zordan
Computer Science Department
University of Southern California
941 West 37th Place, Mailcode 0781
Los Angeles, CA 90089-0781
mataric@cs.usc.edu

Abstract

Humanoid robots will increasingly become a part of human everyday lives, but require more natural and simplified methods for control and human-robot interaction. Our approach to addressing these challenges is to use biologically inspired notions of behavior-based control, and endow robots with the ability to imitate, so that they can be programmed and interacted with through demonstration and imitation. Our approach to the problem, based on neuroscience evidence, structures the motor system into a collection of primitives, which are then used to both generate the humanoid's movement repertoire and provide prediction and classification capabilities for visual perception and interpretation. Thus, what the humanoid can do helps it understand what it sees, and vice versa. We describe the behavior-based background of our work and the neuroscience evidence on which our humanoid motor control and imitation model is based. Next we describe our use of human movement data as input and the humanoid simulation test-bed for evaluation. We follow with a detailed discussion of three means of deriving primitives, the key component of our model, and describe implementations for each of them, as well as experimental results, demonstrated using human movement, captured with vision or magnetic markers, and imitated on a humanoid torso with dynamics, performing various movements from dance and athletics.

Motivation

Humanoid robots will increasingly become a part of human everyday lives, as they are introduced as caretakers for the elderly and disabled, assistants in surgery and rehabilitation, and educational toys for children. To make this possible, the process of robot programming and control must be simplified, and human-robot interaction must be made more natural. Our approach to addressing these challenges is to use biologically inspired notions of behavior-based control, and endow robots with the ability to imitate, so that they can be programmed and interacted with through human demonstration, a natural human-humanoid interface.

Human ability to imitate, i.e., to observe behaviors performed by a teacher and then repeat them, is a poorly understood but powerful form of skill learning. Two fundamental problems in imitation are: 1) segmenting and interpreting the observed behavior, and 2) integrating the visual perception and movement control systems in order to reconstruct what was observed. Our approach to the problem combines the two aspects of the problem into a unified solution based on neuroscience evidence (Matarić 2000, Rizzolatti, Gadiga, Gallese & Fogassi 1996). We structure the motor system into a collection of movement primitives, which are then used to both generate the humanoid's movement repertoire *and* to provide prediction and classification capabilities for visual perception and interpretation of movement. In this way, what the humanoid is capable of **doing** helps it understand what it is **seeing**, and vice versa. The more it sees, the more it learns to do, and thus the better it gets at understanding what it sees for further learning; this is the imitation process.

In this paper we describe the behavior-based background of our work, then provide the neuroscience evidence on which our humanoid motor control and imitation model is based. Next we describe our use of human movement data as input and the humanoid simulation test-bed for evaluation. We follow with a detailed discussion of three means of deriving primitives, the key component of our model, and describe implementations for each of them, as well as experimental results, demonstrated using human movement, captured with vision or magnetic markers, and imitated on a humanoid torso with dynamics, performing various movements from dance and athletics.

Behavior-Based Robotics

Our work in the last decade and a half has focused on developing distributed behavior-based methods for control of groups of mobile robots, and most recently, humanoid agents. Behavior-based control involves the design of control systems that consist of a collection of *behaviors* (Arkin 1998). Behaviors are real-time processes that take inputs from sensors (such as vision, sonar, infra-red), or other behaviors, and send output commands to effectors (wheels, motors, arms), or to other behaviors in the system. The controller, then, is a distributed network of such communicating, concurrently executed behaviors, typically with excellent real-time and scaling properties. The interaction of the behaviors through the environment results in the desired overall system performance.

The inspiration for behavior-based control comes from biology, where natural systems are believed to be similarly organized, starting with spinal reflex movements (Bizzi, Mussa-Ivaldi & Giszter 1991), up to more complex behaviors such as flocking and foraging (Matarić 1995). We have focused on applying the principles of behavior organization to high-dimensional behavior-based systems such as those involved in the control of groups of interacting robots and of humanoids. In both problem domains, we have used basis behaviors, or primitives, described next, in order to structure and simplify the control problem, as well as to enable adaptation and learning.

Basis Behaviors and Primitives

Several methods for principled behavior design and coordination have been proposed (Arkin 1998). In 1992, we introduced the concept of *basis behaviors*, a small set of necessary and sufficient behaviors that could be composed (by sequencing or superposition), as a means of handling controller complexity and simplifying robot programming. Basis behaviors are the *primitives* that serve as a substrate for control, representation, and learning in our behavior-based systems. We first demonstrated their effectiveness on groups of mobile robots. A basis set consisting of avoidance, following, homing, aggregation, and dispersion was used to demonstrate higher-level group behaviors including flocking, foraging/collection, and herding (Matarić 1995). We also demonstrated how, given such a basis behavior set, a learning algorithm could be applied to improve behavior selection over time.

Collections of behaviors are a natural representation for controlling collections of robots. But what is an effective way to use the same idea in the domain of humanoid control, where the individual degrees of freedom of the body are more coupled and constrained? We demonstrate how we have combined the notion of primitives with another line of evidence from neuroscience, that of *mirror neurons* (Rizzolatti et al. 1996), in order to structure humanoid motor control into a general and robust system capable of a variety of skills and learning by imitation.

Humanoid Control and Imitation

Robot control is a complex problem, involving sensory and effector limitations and uncertainty. The more complex the system to be controlled, the more necessary it is to modularize the approach in order to make control viable and efficient. Humanoid agents and robots are highly complex; a human arm has 7 degrees of freedom (DOF), the hand 23, and the control of an actuated human spine is beyond current consideration. Yet humans display complex dynamic behaviors in real time, and learn various motor skills throughout life, often through imitation.

Methods for automating the process of robot programming are in high demand. Reinforcement learning, which enables a robot to improve its behavior based on feedback received from trial and error, are very popular. However, reinforcement learning is slow, as the robot must try various behaviors in different situations repeatedly. Additionally, it can be dangerous to the robot. In contrast, learning by imitation is particularly appealing because it allows the designer to specify entire behaviors by demonstration, instead of using low-level programming or trial and error by the robot. In biological systems, imitation appears to be a complex learning mechanism that involves an intricate interaction between visual perception and motor control, both of which are complex in themselves. Although simple mimicry is found in various animals, so-called "true imitation", involving the ability to learn arbitrary new skills by observation, is only ascribed to very few species, including humans, chimps, and dolphins (Byrne & Russon 1998). This suggests a complex mechanism, which has been implicated as the basis for gestural communication and even language evolution (Arbib 2000). The basis for such a mechanism appears to lie in an evolutionarily older system that combines perception and motor control, and enables mimicry, described next.

Neuroscience Inspiration

Evidence from neuroscience studies in animals points to two neural structures we find of key relevance to imitation: spinal fields and mirror neurons. Spinal fields, found in frogs and rats so far, code for complete primitive movements (or behaviors), such as reaching and wiping (Bizzi et al. 1991). More interestingly, they are additive; when multiple fields are stimulated, the resulting movement is a meaningful combination. Since the spine codes a finite number of such fields, they represent a basis set of primitives, and were precisely the inspiration for our work on basis behaviors, described above.

Neurons with so-called "mirror" properties were recently found in monkeys and humans. They appear to directly connect the visual and motor control systems by mapping observed behaviors, such as reaching and grasping, to motor structures that produce them (Rizzolatti et al. 1996). It is not yet known how many such movements are directly mapped by the mirror system, but the basic idea serves as rich inspiration for structuring a robotic imitation system.

We combine these two lines of evidence, spinal basis fields and mirror neurons, into a more sophisticated notion of behaviors, or *perceptual-motor primitives* (Matarić 2000). These allow a complex system, such as a humanoid, to recognize, reproduce, and learn motor skills. As mentioned above, the primitives are used as the basis set for generating movements, but also as a "vocabulary" for classifying observed movements into executable ones. Thus, primitives can classify, predict, and act.

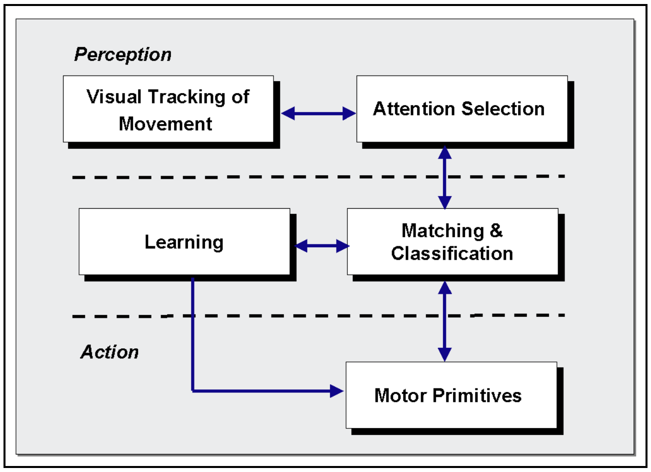

Figure 1: *Our imitation architecture, structured around perceptual-motor primitives used for movement segmentation, classification, and generation.*

In our approach to imitation, the vision system continually matches any observed human movements onto its own set of motor primitives. The primitive, or combination of primitives, that best approximates the observed input also provides the best predictor of what is expected to be observed next. This expectation facilitates visual segmentation and interpretation of the observed movement. Imitation, then, is a process of matching, classification, and prediction, and learning by imitation, in turn, is the process of creating new skills as novel sequences and superpositions of the matched and classified primitives. Our imitation approach is hierarchical in structure; it allows the robot to initially observe and imitate a skill, then perfect it through repetition, so that the skill becomes a routine and thus a primitive itself. As a result, the set of primitives can be adapted over time, to allow for learning arbitrary new skills, i.e., for "true" imitation.

Figure 1 shows the overall structure of our imitation architecture, including the visual perception and attention module, the classification module, and the motor primitives (Matarić 2000, Jenkins, Matarić & Weber 2000). The learning component is also shown, allowing adaptation both at the level of classification, for finding a closer match to the observed behavior, and repetition for optimizing and smoothing the performance.

Before going on to discuss in more detail the notion of primitives and our implementation and validation, we first describe the means by which we acquire training as well as evaluation data for our model.

Using Human Movement Data

We use human movement data to derive the primitives, as well as to test the implementations of our model.

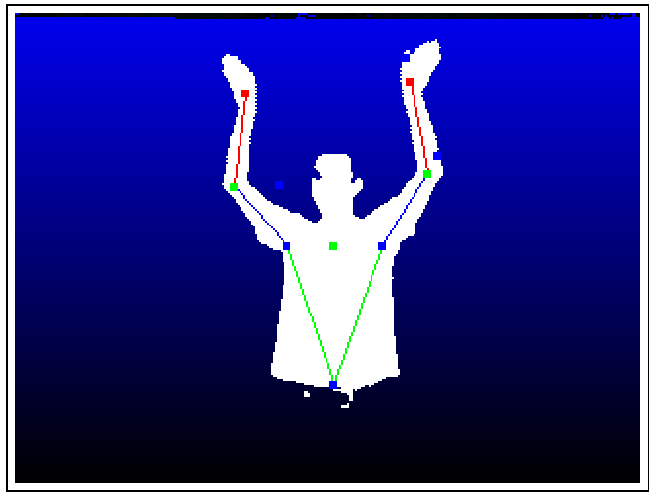

Figure 2: *A snapshot of the output of our vision-based motion tracking system.*

Different types of data are used, including visual motion tracking, Cartesian magnetic markers, and joint angle data. We have developed a motion tracking system for extracting features from a video stream (Weber 2000). This system is simple and can be replaced by more sophisticated commercial versions, but is capable of selecting a collection of features from the moving image, based on a constrained (un-occluded and unambiguous) initial position and kinematic model of the object/body being imitated. This greatly simplifies finding the initial match between the features in the visual image and the body being observed. The match enables tracking of the body over time, and allows for fast computation and updating of current limb position, and simple prediction of future position, used in turn to speed up recognition. Our ongoing work will address how the primitives themselves also provide further predictive capability for the visual tracking. Figure 2 shows a snapshot of the output of the motion tracking system.

We are also using 3D magnetic marker data from the human arm, gathered from subjects imitating videos of arm movements while wearing FastTrak markers for position recording[1] We used four markers: near the shoulder, the elbow, the wrist, and the start of the middle finger. The movement data resulting form this experiment are being used as input into our imitation system, as well as for automatically learning primitives. Methods for analyzing this type of data as well as the associated psychophysical results are reported in (Pomplun & Matarić 2000).

Finally, we are using full-body joint angle data gathered with the Sarcos SenSuit, a wearable exoskeleton that simultaneously records the joint positions of 35

[1]These data were gathered at the National Institutes of Health Resource for the Study of Neural Models of Behavior at the University of Rochester.

degrees of freedom: the shoulders, elbows, wrists, hips, knees, ankles, and waist.[2] We are currently focusing on upper-body movements, and reproducing those on the test-beds described next.

Evaluation Test-beds

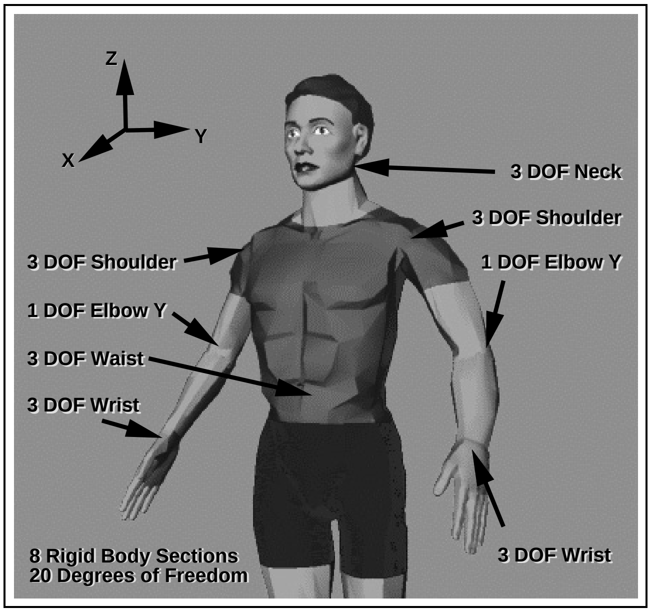

Figure 3: *The Adonis 20 DOF full-body dynamic simulation.*

We are conducting a large number of experimental trials to properly validate our approach to humanoid motor control and imitation. Most of our work so far has been done on Adonis, a 3-D rigid-body simulation of a human, with full dynamics (Figure 3). Mass and moment-of-inertia information is generated from the graphical body and human density estimates. Equations of motion are calculated using a commercial solver, SD/Fast. The simulation acts under gravity and accepts external forces from the environment. The static ground removes the need for explicit balance control. We have added joint limits and self-collision detection. Adonis consists of eight rigid links connected with revolute joints of one DOF and three DOF, totaling 20 DOF, but most of our experiments use the 13 DOF of Adonis' upper body, and currently do not address locomotion.

As the project progresses, we hope to apply our approach to the NASA Robonaut humanoid upper body, as well as the Sarcos full-body humanoid robot[3] as the ultimate, physical, test-beds for humanoid behavior.

[2]These data are obtained trough a collaboration with the ATR Dynamic Brain Project at the Human Information Processing Labs in Kyoto, Japan.

[3]Also through a collaboration with the ATR Dynamic Brain Project at the Human Information Processing Labs in Kyoto, Japan

Visual Classification Into Primitives

Behavior primitives are the unifying mechanism between visual perception and motor control in our approach. Visual perception is an important constraint on the primitives, and a key component of the imitation process. Since the human (and humanoid) visual attention is resource-limited, it must select the visual features that are most relevant to the given imitation task. Determining what those features are for a given demonstration is a challenging problem.

Our previous work showed that people watching videos of arm movements show no difference in attention whether they are just watching, or intending to subsequently imitate. In both cases, they fixate at the end-point, i.e., the hand or a held object (Matarić & Pomplun 1998). However, it is impossible to classify all possible end-point trajectories into a useful set of task-independent categories or primitives. Fortunately, this is not necessary, since the imitation system is targeted for mapping observed movement of bodies similar to the observer's, to the observer's own motor repertoire. Accordingly, the mirror system is sensitive to biological motion of similar bodies (for example, in monkeys it was shown to respond to monkey and human movements). Furthermore, although human(oid) bodies are potentially capable of vast movement repertoires, the typical, everyday movement spectrum is not nearly as large.

Consequently, the visual perception mechanism can be effectively biased toward recognizing movements it is capable of executing, and especially those movements it performs most frequently. The structure of the motor control system, and its underlying set of movement primitives, provides key constraints for visual movement recognition and classification.

Choosing the Primitives

Choosing the right primitives is a research challenge, driven by several constraints. On the one hand, the motor control system imposes physical bottom-up limitations, based on its kinematic and dynamic properties. It also provides top-down constraints from the type of movements the system is expected to perform, since the primitives must be sufficient for the robot's entire movement repertoire. On the other, the choice of primitives is also influenced by the structure and inputs into the visual system, in order to map the various observed movements into its own executable repertoire.

In order to serve as a general and parsimonious basis set, the primitives encode groups or classes of stereotypical movements, invariant to exact position, rate of motion, size, and perspective. Thus, they represent the "generic" building blocks of motion that can be implemented as parametric motor controllers. Consider for example a primitive for reaching. Its most important parameter is the goal position of the end-point, i.e., hand or held object. It may be further parametrized by a default posture for the entire arm. Such a primitive enables a robot to reach toward various goals within a multitude of tasks, from grasping objects and tools, to dancing, to writing and drawing. We used just such a reaching primitive in our experiments, in order to, for example, reconstruct the popular dance Macarena (Matarić, Zordan & Williamson 1999).

This approach to motor control stands in sharp contrast to the explicit planning approach to controlling robot manipulators, which computes trajectories at run-time, whenever they are needed. While fully general, on-demand trajectory generation is computationally expensive and potentially slow. In our approach, instead of computing trajectories *de novo*, stereotypical trajectories are built-in as well as learned, then merely looked up and parameterized for the specific task at hand. The notion of primitives takes advantage of the fact that it is simpler to learn and reuse an approximation of the inverse kinematics for specific areas in the workspace or a specific trajectory, than it is to compute them anew each time.

What constitutes a good set of primitives? We have experimented with three means of obtaining this set: 1) designing and encoding the set by hand (Matarić et al. 1999), 2) using human movement data to extract the primitives by hand (Jenkins et al. 2000), and 3) using human movement data to automatically extract the primitives (Fod, Matarić & Jenkins 2000). We describe each of the approaches and their results below.

Manually Coded Primitives

In our first approach, we designed the primitives by hand. Initially, we implemented two versions of the spinal fields described above. One closely modeled the frog data, and used a joint-space representation, i.e., it controlled individual joints. The other used another biologically-inspired approach, impedance control (Hogan 1985), which operates in the external coordinate frame, in our case the humanoid's hands. Our impedance motor controller applied forces at the hands and "dragged" the rest of the arm along. We also used a default posture for the arm, which provided natural-appearing whole-arm movements that reached the desired hand destination (Matarić et al. 1999).

We tested both types of primitives on a complicated sequential motor task, dancing the Macarena. Both were shown to be effective, but each had limitations for particular types of movements. This has led us to propose and explore a combination approach, where multiple types of primitives can be sequenced and combined. Specifically, we constructed a basis behavior set consisting of three types of primitives: 1) discrete straight-line movements using impedance control; 2) continuous oscillatory movements using coupled oscillators (or a collection of piece-wise linear segments using impedance control); and 3) postures, using PD-servos to directly control the joints. We also added a forth type of primitive, for avoidance, implemented as a repulsive vector field. The fourth primitive was continuously active, and combined with whatever other primitive was being executed, in order to prevent any collisions between body parts. In the Macarena, for example, this is necessary for arm movements around and behind the head (Figure 4).

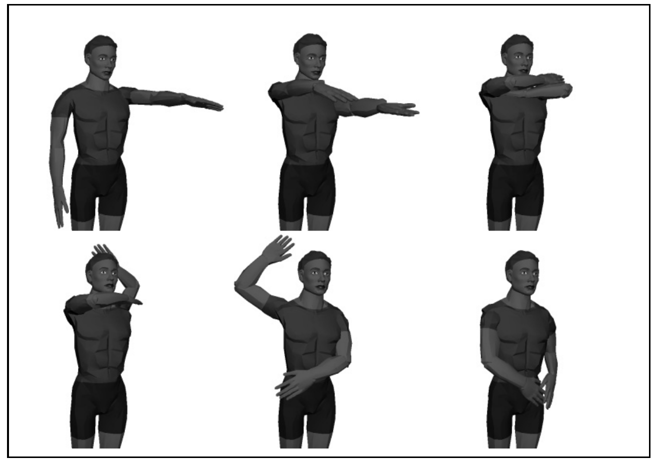

Figure 4: *Snapshots of Adonis dancing the Macarena.*

Manually Extracted Primitives

In the second approach, we used a motion sequence of a human executing a small set of movements, automatically extracted features from those, and then manually extracted a set of primitives from that movement sequence. The features were extracting using the vision-based motion tracking system described above (Weber 2000), and consisted of the hands, elbows, shoulders, and the center of the torso. These were then manually segmented in time to isolate each execution of a primitive. The primitives used in this particular implementation were based on the set of 2D geometries, and included lines, circles, and arcs. Each segment of motion by the human performer was used as the description of a motion generated from a certain primitive and was converted into a representation amenable for classification. Thus, a wide variety of motions could be classified into the categories of the "training" set. Once these primitives were established, any observed movements was described with only the trajectories of the end-points (i.e., hands). The trajectories were represented as normalized gradients, which are invariant to scaling and translation.

Our primitive classifier uses the descriptions of the primitives to segment a given motion based on the movement data. In these experiments (Weber, Jenkins & Matarić 2000, Jenkins et al. 2000), end-point data for both arms were used as input for the vector quantization-based classifier (Arya & Mount 1993). As discussed, a key issue in classification is representing the primitives such that they account for significant invariances, such as position, rotation, and scaling. In our approach to classification, the original motion is formed into a vector of relative end-point movements between successive frames, then smoothed and normalized. All other information about the movement, such as global position and arm configuration, is ignored at the classification level. This allows for a small set of high-level primitive representations instead of a potentially prohibitively large set of detailed ones. Instead, the details of the observed movement can be used for parameterizing the selected primitive(s) at the level of movement reconstruction and execution.

To simplify matching, primitives themselves are described in the same normalized form. For each timestep of the observed motion, a fixed-horizon window is compared to every primitive and the one that matches the input best is selected. Adjacent windows with identical classifications are connected to form continuous segments. For any segments that fail to match any of the given primitives, the reaching primitive is used to move the end-point in a frame-by-frame manner. Because the horizon window is of fixed size, the perception of a distinct match of a primitive applies only for the given time-scale. We are currently working on addressing classification at multiple time-scales. More details about the classification algorithm can be found in Weber et al. (2000).

To validate this approach, we used Adonis to demonstrate various imitation tasks, including reaching, ball throwing, aerobics moves, and dance. Figure 5 summarizes the implementation of the imitation system, including the visually extracted human movement features, and the validation on Adonis. The next section discusses the imitation results in more detail.

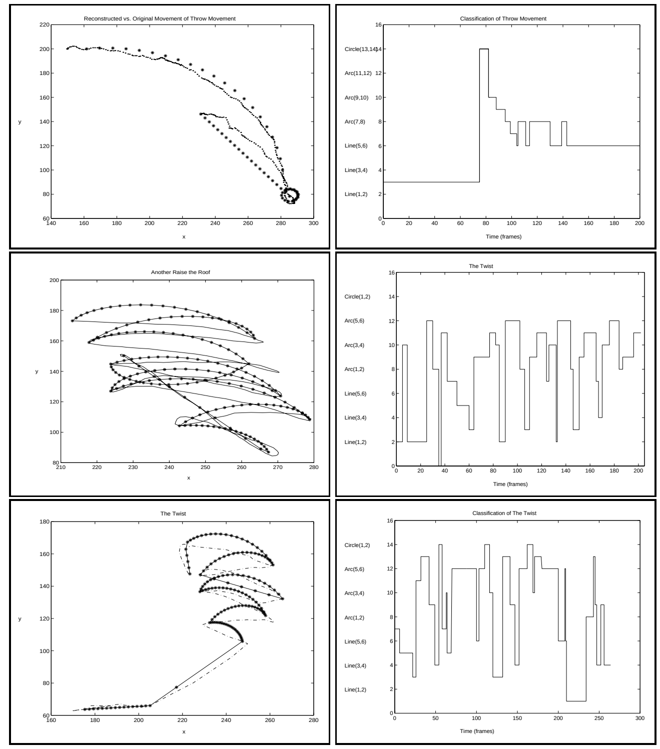

Figure 6: *A selection of imitation results showing, from the top down, a throw, twists, and arm raises. On the left, plots of the human demonstrated movements and reconstructed motions; on the right, the classification results for each.*

Results of Movement Reconstruction

Primitives provide higher-level descriptions of a movement and visually observed metric information is used to parameterize those descriptions for generating executable movement. This allows for a small number of general primitives to represent a large class of different movements, such as reaches to various destinations on and around the body. Note that it is not our goal to achieve perfect, completely precise, high-fidelity imitation. While that may be possible through the use of exact quantitative measurements of the observed movement using signal processing techniques, it is not what happens in imitation in nature, and it is neither necessary nor helpful for our main goals: natural interaction and programming of robots. For those purposes, we aim for an approximation of the observed behavior, one that allows any necessary freedom of interpretation by the humanoid robot, but achieves the task and/or effectively communicates and interacts with the human. Our goal is also distinct from "task-level" imitation, which only achieves the goal of the demonstration, but does not imitate the behaviors involved. This problem has been studied in assembly robotics; where a series of visual images of a human performing an object stacking task was recorded, segmented, interpreted, and then repeated by a robotic arm (Kuniyoshi, Inaba & Inoue 1994). While the focus of that work was on visual segmentation into sub-goals, we aim for a more biologically-motivated model of imitation: a system capable of imitating the observed behavior, i.e., the process that brings the sub-goals about, with sufficient, but not perfect, accuracy.

Since the movement primitives represent whole-arm movements, our visual system can use either only the end-point (hand) location for imitation, or can gather higher fidelity data that includes the position of the elbow and other markers. In the experiments described here (and in more detail in Weber et al. (2000) and Jenkins et al. (2000)), the position of the hand over time was sufficient for effective imitation. The hand trajectory over time is segmented using the classifier, which, at each point in time, matches the expected output of each of the primitives with the observed input, and selects the best match. Consecutive matches of the same primitive indicate a higher confidence in the match. The output of the classification is a sequence of primitives and their associated parameters. These are then sent to the motor control system, and activate the primitives in turn, in order to reconstruct the observed behavior.

We selected imitation tasks from athletics, danc-

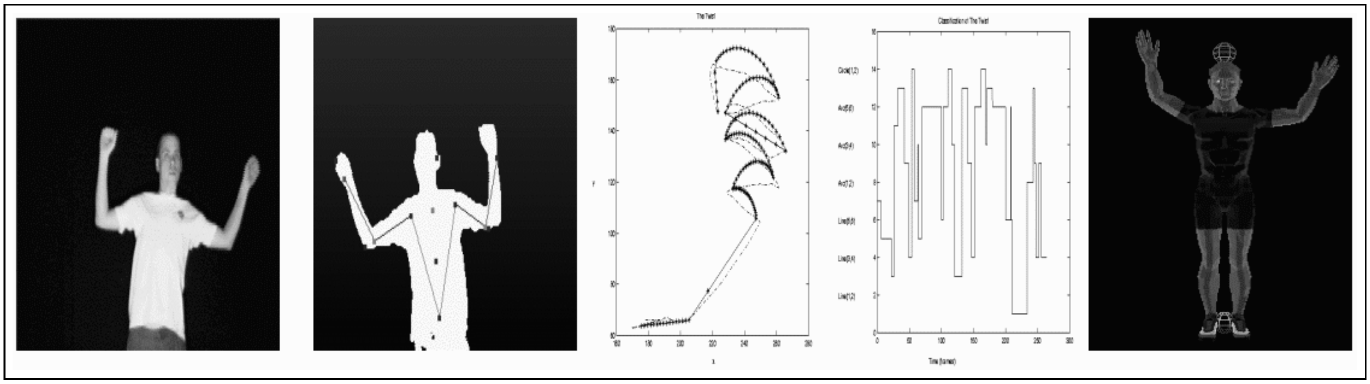

Figure 5: *The imitation process using manually extracted primitives. The human demonstrator is shown on the far left, followed by the visually-extracted features, a plot of the demonstrated and imitated trajectories, the classification results, and finally the imitation as performed by Adonis.*

ing, and aerobics; 6 tasks were performed in front of the vision system and presented to our imitation system. Figure 6 shows an example of the imitation system performance on three human-demonstrated movements: a throw, a twist, and a raising of the arms. Videos demonstrating imitation of various types of movements, including details of the input, tracking system, associated mapped primitives, and Adonis imitating can be found at http://www-robotics.usc.edu/∼ agents/imitation.html. We are currently actively pursuing methods for evaluating the quality of the resulting imitation (Pomplun & Matarić 2000).

Automatically Derived Primitives

We have developed a method to automatically organize and derive the structural elements or primitives directly from movement data. As in biological systems, these primitives can be superimposed and sequenced. We hypothesize that the trajectory executed by the arm is composed of segments in which a set of principal components are active in varying degrees. We first segment movement data and then apply principal component analysis (PCA) on the resulting segments to obtain "eigen-movements" or primitives. The eigenvectors corresponding to a few of the highest eigenvalues provide us with a basis set for a subspace. The projection of the segment vector onto this subspace contains most of the information about the original segment. By clustering in this subspace we obtain a set of points that correspond to a set of frequently used movements which can be used to calibrate controllers. To evaluate the method of movement encoding in terms of eigenmovement primitives, and the subsequent reconstruction of the original movements, we calculated the mean square deviation of the reconstructed movement. Finally, we demonstrate the movements on a dynamic humanoid simulation.

Figure 7 summarizes our approach and the methods

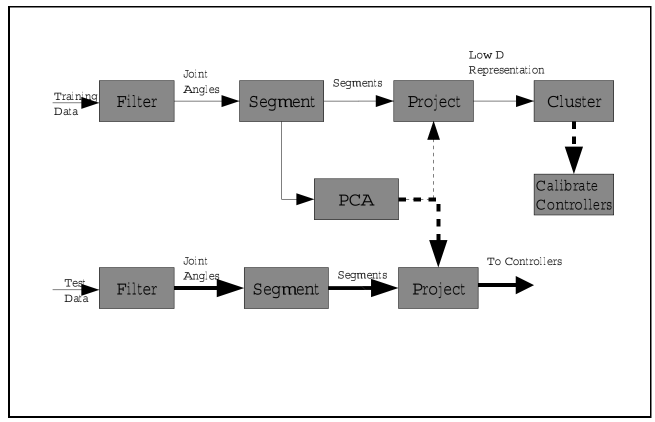

Figure 7: *Components of the primitive deriving system. Thin lines represent the flow of data in the training phase. Dark lines represent the flow of data for the testing phase. Dotted lines represent the use of previously computed data.*

involved. Human data is fed into the system as a temporal sequence of vectors. We chose to use a joint angle representation of data in our analysis. Consequently, in the preprocessing step, we used inverse kinematics to transform 3D marker coordinates into a joint angle representation. We then filter the data to remove the incidental random zero-velocity points caused by noise. Next, we segment the movement data so that a set of finite dimensional vectors could be extracted. Using PCA, we obtained a set of eigenvectors, a few of which represent almost all the variance in the data. We then applied K-means clustering to group the projections of the segments in order to obtain clusters that represent often used movements. These clusters are designed to provide calibration points for the control system we design for the system.

To validate and evaluate the performance of the above representation, we projected and reconstructed

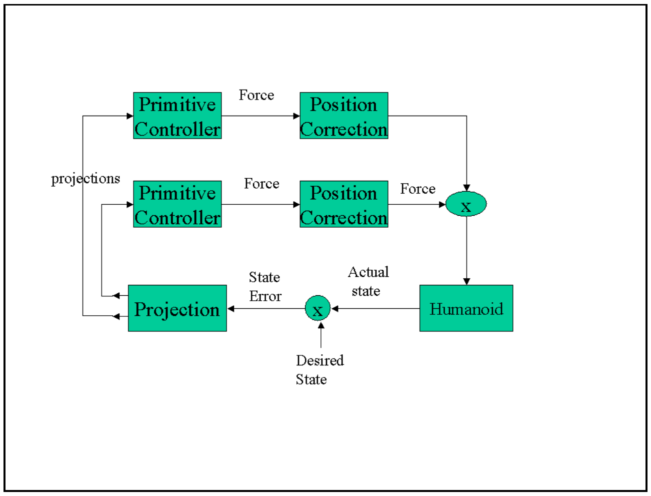

Figure 8: The model of the controllers. Each primitive has an individual Primitive Controller. The Position Correction module adjusts to changes in the location of the center of the stroke. In this work, the Humanoid is modeled by Adonis. The Projection Module projects the errors back onto the eigenmovement space.

some test data. We first perform filtering and segmentation and then project the data on to the eigenvectors we had previously derived. The reconstruction involves summing up the component eigenvectors weighted by their respective coefficients. A detailed discussion can be found in (Fod et al. 2000).

Our continuing research focuses on methods for generating controllers for the automatically derived primitives. One model we are exploring is shown in Figure 8. Desired inputs are fed into the control system and the error is calculated, then projected onto the primitives derived above. We thus obtain certain projection coefficients which generate the desired control inputs for the primitive controllers. As shown in the figure, each of the primitive controllers executes an individual primitive. Given the projection coefficients, the controllers generate appropriate force signals. The position correction modules correct for the difference in the center of the stroke and the consequent change in the dynamics of the humanoid. The force signals are then sent to Adonis. The resulting angle measurements are compared with the desired angles and correction signals are sent to the primitive controllers.

We are also working on the policies needed to modify these controllers for changes in duration and couplings with other controllers that are simultaneously active. This is an important part of our model, which involves simultaneous execution as well as sequencing of the primitives for both movement perception and generation (Matarić 2000).

Conclusions

Control and interaction with humanoids is a complex problem. We believe that imitation provides a promising means of addressing both problems by presenting a natural human-humanoid interface. It allows humanoid robots to become accessible to human operators with various levels of expertise. This, in turn, can be used to make humanoids more useful in human environments for tasks ranging from the highly technical (e.g., surgery and object assembly), to the more interactive dynamic ones (e.g., education, care taking, and entertainment).

Given the complexity of the humanoid control and imitation problems, we turned to biology for inspiration and developed an approach based on current neuroscience theories of how motor control and imitation may be organized. However, our approach is firmly grounded in already validated principles of scalable and robust behavior-based control. Our results to date demonstrate the effectiveness of the basis behavior primitive approach for both motor control and imitation. Additional information and videos of our systems can be found at: http://www-robotics.usc.edu/~agents/imitation.html

Acknowledgements

This work is supported by the NSF Career Grant IRI-9624237 to Maja Matarić and the USC All-University Predoctoral Fellowship to Chad Jenkins. The authors are grateful to Prof. Dana Ballard and Prof. Mary Hayhoe for making the resources available for performing psychophysical experiments at the NIH Resource for the Study of Neural Models of Behavior at the University of Rochester.

References

Arbib, M. (2000), The Mirror System, Imitation, and the Evolution of Language, *in* K. Dautenhahn & C. Nehaniv, eds, 'Imitation in Natural and Artificial Systems', MIT Press.

Arkin, R. C. (1998), *Behavior-Based Robotics*, MIT Press, California.

Arya, S. & Mount, D. M. (1993), Algorithms for fast vector quantization, *in* J. A. Storer & M. Cohn, eds, 'Proceedings of the Data Compression Conference', IEEE Press, pp. 381–390.

Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. (1991), 'Computations Underlying the Execution of Movement: A Biological Perspective', *Science* **253**, 287–291.

Byrne, R. W. & Russon, A. E. (1998), 'Learning by Imitation: a Hierarchical Approach', *The Journal of Behavioral and Brain Sciences* **16**, 3.

Fod, A., Matarić, M. J. & Jenkins, O. C. (2000), Automated Derivation of Primitives for Movement Classification, *in* 'Proceedings, First IEEE-RAS International Conference on Humanoid Robotics', Cambridge, MA, MIT.

Hogan, N. (1985), 'Impedance Control: An Approach to Manipulation', *Journal of Dynamic Systems, Measurement, and Control* **107**, 1–24.

Jenkins, O. C., Matarić, M. J. & Weber, S. (2000), Primitive-Based Movement Classification for Humanoid Imitation, *in* 'Proceedings, First IEEE-RAS International Conference on Humanoid Robotics', Cambridge, MA, MIT.

Kuniyoshi, Y., Inaba, M. & Inoue, H. (1994), 'Learning by watching: extracting reusable task knowledge from visual observation of human performance', *IEEE Transactions on Robotics and Automation* **10**(6), 799–822.

Matarić, M. J. (1995), 'Designing and Understanding Adaptive Group Behavior', *Adaptive Behavior* **4**(1), 50–81.

Matarić, M. J. (2000), 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.

Matarić, M. J. & Pomplun, M. (1998), 'Fixation Behavior in Observation and Imitation of Human Movement', *Cognitive Brain Research* **7**(2), 191–202.

Matarić, M. J., Zordan, V. B. & Williamson, M. (1999), 'Making Complex Articulated Agents Dance: an analysis of control methods drawn from robotics, animation, and biology', *Autonomous Agents and Multi-Agent Systems* **2**(1), 23–44.

Pomplun, M. & Matarić, M. J. (2000), Evaluation Metrics and Results of Human Arm Movement Imitation, *in* 'Proceedings, First IEEE-RAS International Conference on Humanoid Robotics', Cambridge, MA, MIT.

Rizzolatti, G., Gadiga, L., Gallese, V. & Fogassi, L. (1996), 'Premotor cortex and the recognition of motor actions', *Cognitive Brain Research* **3**, 131–141.

Weber, S. (2000), Simple Human Torso Tracking from Video, Technical Report IRIS-00-380, University of Southern California, Institute for Robotics and Intelligent Systems.

Weber, S., Jenkins, O. C. & Matarić, M. J. (2000), Imitation Using Perceptual and Motor Primitives, *in* 'Proceedings, The Fourth International Conference on Autonomous Agents (Agents 2000)', Barcelona, Spain, pp. 136–137.