Puppet Master: Designing Reactive Character Behavior by Demonstration

James E. Young*
University of Calgary, JST ERATO

Takeo Igarashi†
University of Tokyo, JST ERATO

Ehud Sharlin‡
University of Calgary

Abstract

We present Puppet Master, a system that enables designers to rapidly create interactive and autonomous character behaviors (e.g., of a virtual character or a robot) that react to a main character controlled by an end-user. The behavior is designed by demonstration, allowing non-technical artists to intuitively design the style, personality, and emotion of the character, traits which are very difficult to design using conventional programming approaches. During training, designers demonstrate paired behavior between the main and reacting characters. During run time, the end user controls the main character and the system synthesizes the motion of the reacting character using the given training data. The algorithm is an extension of image analogies [Hertzmann et al. 2001], modified to synthesize dynamic character behavior instead of an image. We introduce non-trivial extensions to the algorithm such as our selection of features, dynamic balancing between similarity metrics, and separate treatment of path trajectory and high-frequency motion texture. We implemented a prototype system using physical pucks tracked by a motion-capture system and conducted a user study demonstrating that novice users can easily and successfully design character personality and emotion using our system and that the resulting behaviors are meaningful and engaging.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques;

Keywords: Demonstration, Artistic Design, Interactive Animation, Interaction

1 Introduction

Characters, such as those in animated films or computer games, or even autonomous robots interacting in the real world, are becoming increasingly common in everyday life. Having convincing, believable personalities and behaviors is very important for these characters, as it strengthens communication and suspension of disbelief, and ultimately results in a more rewarding, engaging, and comfortable experience [Bates 1994; Breazeal 2002; Reeves and Nass 1996]. In particular, it is critically important that interactive characters react convincingly to real-time user input while maintaining a coherent personality. For example, an aggressive merchant in a video game may chase after the user"s character, a shy cleaning robot may hide from humans while cleaning, and a pet robot dog may jump happily when its owner returns home.

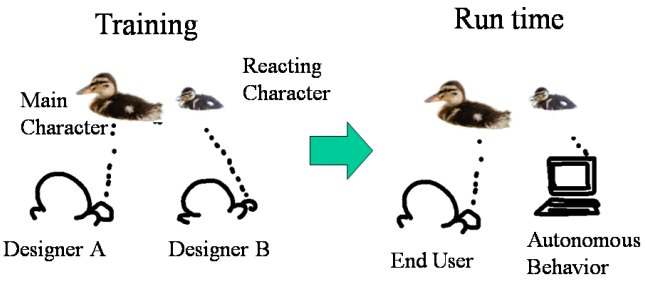


Figure 1: *During training, the system learns the behavior of a reacting character in response to a main character's behavior. At run time, the system synthesizes a reacting character's behavior, reproducing the personality and emotion demonstrated in the training.*

Programming a character's real-time interactive behavior is a difficult problem, particularly when trying to achieve a certain personality or interaction style. The logical, step-by-step state-machine style endorsed by conventional programming languages is good for specifying goal-oriented actions, but is not well suited to the design of human-like traits such as personality and emotion; consider how difficult it is to program shyness or anger using a series of logical states, rules and steps.

Artists, however, such as actors, puppeteers, and so forth, have an incredible ability to develop interactive characters and personalities in various performance-based mediums such as puppet shows, plays, computer avatars and even remote-controlled robots. Unfortunately, conventional tools available to create autonomous interactive behaviors are often inaccessible to non-technical artists, in part because of their algorithmic nature and their incapability to explicitly express character personality and emotion. The result is that interactive behaviors are generated primarily by computer software engineers using logic-based algorithms, with results often being predictable and boring.

In this paper, we introduce a programming-by-demonstration approach to make the design of interactive character behavior accessible to artists. In the training phase a user (or two working collaboratively) demonstrates paired motion of two characters, with one character reacting to the actions of the other. At run-time, the end-user controls the main character and the system generates, in real time, the reactive behavior of the other character with the characteristics observed in the training data (Figure 1). Demonstrating an interactive behavior allows artists to encapsulate personality and emotion that they may not be able to logically explain using computer algorithms. Furthermore, training is quick (average 33s in our study) and generation is done in real time without preprocessing.

Our behavior synthesis algorithm is an extension of the image analogies algorithm [Hertzmann et al. 2001], which learns static image filters from example image pairs and applies them to a new input image. Similarly, our system learns reactive behavior from an example pair of motions and applies it to a new input motion. Our paper describes how we extend the original method to work for real-time, dynamic, reactionary locomotion behavior. Specifically, we introduce meaningful behavior-related features, a method

*e-mail: jim.young@ucalgary.ca
†takeo@acm.org
‡ehud@cpsc.ucalgary.ca

for balancing between the similarity and coherence metrics, and we also separately synthesize general motion trajectory and motion texture, integrating them during the final stages of motion synthesis.

We built two prototype systems, one using a standard mouse as input and the other using a tabletop system with physical pucks tracked by a motion capture system. The mouse-based system allows the user to control the movement of one character at a time, while the tabletop system allows for simultaneous control of both the main and reacting characters, and allows the user to control the orientation of the two characters. We ran a user study using the tabletop system asking one group of participants to design a set of character behaviors and the other group to interact with a set of designed behaviors. The results shows that novice users can design interactive behaviors quickly using our system and successfully convey personality and emotion in the form of interactive behavior.

2 Related Work

There has been a great deal of work that aims for life-like, convincing interactive behavior. One common approach is through explicitly programming the behavior model [Blumberg and Galyean 1995; Maes 1995; Reynolds 1987], where the programmer explicitly and directly defines what to do for particular input scenarios. These systems require an understanding of the underlying algorithm and so are less accessible to the general artist, and do not support direct and intuitive design of emotion, personality, and style.

Programming by demonstration was originally used to automate GUI operations [Cypher 1991; Maulsby et al. 1989], e.g., Pavlov [Wolber 1997], explicitly for interactive agents, defines the low-level stimulus-response behavior of the agent. These systems define behavior using logical event sequences and conditionals, and do not provide tools to represent personality and emotion.

Several systems design animation by performance demonstration [Dontcheva et al. 2003; Hertzmann et al. 2002; Igarashi et al. 2005b; Igarashi et al. 2005a; Thorne et al. 2004], and some apply the idea to robotic motion [Frei et al. 2000; Raffle et al. 2004]. These systems, however, focus on the playback of the demonstration and do not respond intelligently to user interaction.

Other systems focus on synthesizing new motion from a large, preprocessed motion example database in real time [Lee and Lee 2004; Lerner et al. 2007; Wiley and Hahn 1997]. While some systems interactively respond to user input (joystick control, moving obstacles, other characters, etc), the mapping from the user input to the output is explicitly (and often tediously) defined by the programmer. Furthermore, the target of these systems is primarily the plausibility of physical motion (punch, jump, walk, collision avoidance, etc.), not the explicit design of character or personality emerging from interactive motion.

Research in human-robot interaction is shifting attention from viewing robots as tools or laborers, to designing affective and sociable robotic interfaces [Breazeal 2002; Norman 2004]. Robots are active participants in their users' physical environments, and as such the design of their form, posture and movement play a dramatic role in the quality of the resulting interaction [Matsui et al. 2005]. Designing robotic actions and movement by demonstration has been investigated extensively (for example, [Breazeal 2002; Frei et al. 2000; Raffle et al. 2004; Matsui et al. 2005]) with efforts ranging from simple movement playbacks to complex integrated actions. However, again, these methods were targeting goal-oriented design of robotic pose and motion, and did not allow to explicitly convey emotion or personality.

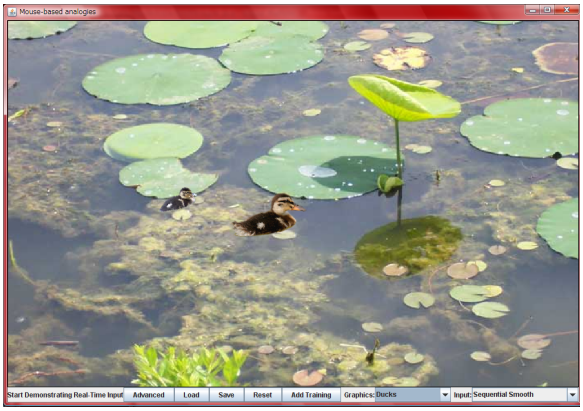


Figure 2: *The mouse GUI. Notice the large work area and lack of parameters and settings.*

3 System Overview

The ultimate goal of our research is to allow the direct and intuitive design of all aspects of reactive behavior (gestures, facial expressions, eye movement, pose, etc) to create believable whole-body characters. As an initial experiment to explore the programming of characters by demonstration, this paper focuses on character locomotion (movement paths). We have implemented two interfaces that enable designers to demonstrate behaviors: a mouse-based GUI and a tabletop Tangible User Interface (TUI).

3.1 Mouse-Based GUI

The mouse-based GUI (Figure 2) is fast and easy to use on any regular desktop PC. The standard mouse, however, lacks a rotation sensor and is designed for use by a single-user. This means that the main character and reactor behaviors must be trained sequentially and that the direction a character is looking cannot be explicitly specified, e.g., a character cannot move sideways or back away. With sequential training, the designer first demonstrates an example input path to represent the main character. Then, the system replays the main-character"s motion path while the designer demonstrates the sample reaction. For behavior generation, the user simply controls the main character and reaction is generated in real time. Here, the entity"s look direction is matched to the movement direction.

Figure 3: *A user interacting with the Vicon TUI system.*

3.2 Tabletop Tangible User Interfaces (TUIs)

Training interactive behaviors by demonstration is reminiscent of acting or puppetry; actions generally done in the physical world, away from traditional computers. Using a mouse mapped to an on-screen avatar removes this direct physicality and forces an artist to map their intentions through the arguably less-than-intuitive mouse interface, separating the artist action and perception space [Fjeld et al. 1986; Sharlin et al. 2004].

Our Tangible User Interfaces (TUIs)-based tabletop design (see Figure3) allows artists to use physical objects (the TUIs) as handles to demonstrate behavior. The TUIs are both input to the system and offer immediate physical feedback of the system state, providing intuitive interaction not possible with traditional interfaces[Ishii and Ullmer 1997]. Using TUIs also allows for character orientation, and multiple TUIs can be used for simultaneous training of both the main and reacting character behaviors, either two-handed by one designer or by a pair of collaborators.

4 Algorithm

During training, the system simply stores the paired motion data. During generation, the algorithm compares the run-time situation between the main and reacting character to the training data. The training most similar to the current situation is used to direct the generation of the next character output; the generation is based on a mix of training data from various source locations. This method avoids large chunks of consecutive, static replay of training data, and reacts immediately to changing user input, while maintaining the style and properties of the training data.

The key problems to be solved are to a) select a similarity metric with meaningful features that matches behavioral similarity as end-users may recognize it, and b) ensure the generation output algorithm maintains the characteristics and textures given in the training data; all of this must happen in real time. We offer solutions to both of these problems.

4.1 Algorithm Formalization

The variables $< I, R >$ represent time-series data for the two characters, where I is the main and R is the reacting character. Following, $< I^t, R^t >$ are the training time-series data used by the generation system, which is simply recorded as it is demonstrated. Pseudo code for the generation is given below: `BestMatch()` finds training data most similar to the current situation, and `Generate()` uses the selected data to direct the next output action. New user input is not used until the subsequent iteration because features used in tracking (such as relative character position) require data for both characters. Our implementation operates at 40Hz so this delay is negligible.

```
Loop
    e = BestMatch(I^t, R^t, I, R)
    newMovement = Generate(e)
    R.append(newMovement)
    I.append(getNewInput())
```

4.2 Data and Features

Our data is a time-dependent array containing the location, x,y, and the direction d of each entity at each time interval. From here we select and extract various features, a core part of our approach that decides what characteristics of behavior will be matched and generated. We generally ignore world-coordinates, focusing on the relationship between the two entities and changes in local state. The similarity metrics use these features over a time window to encapsulate a trend over time. We explored many features not discussed here (e.g., direct path data, distance and delta distance between characters), and settled on the following features, as illustrated in Figure 4. We omit a detailed discussion here on reasons for this final feature selection for brevity. The use of these features is different for each algorithm step as outlined in each respective section.

Velocity – calculated by taking the magnitude of the vector between an entity"s position and its previous one. This captures speed and acceleration-related aspects of behavior such as different reactions for stopped, accelerating, fast, or slow input.

Relative position – position of the reactor in relation to the main character"s position and look direction (coordinate space). This captures 'relational behavior such as following, circling, and approaching. This is two scalar values, one per axis, representing how much the reacting character is behind or in front of, and to the left or right of the main character.

Normalized look direction – look direction normalized to the movement direction, with 0 pointing forward. This captures the relationship between where a character is looking and moving, e.g., if a character is backing up or moving sideways.

Relative look direction – the difference between the entities' look directions. This captures a character turning away shyly when observed or aggressively facing an opponent.

Absolute Movement direction – the vector from the character"s previous world-coordinate position to its current one. This feature, treated separately from the others, does not involve the other entity. The generation phase (Section 4.4) uses this to add high frequency texture to the output motion.

ΔDirection – change in direction from one step to the next, represents the shape of the locomotion path (not in relation to the other entity). This feature helps to identify similar movement shapes and styles such as shaky or smooth.

4.3 BestMatch (Similarity Metric)

Our similarity metric is based on Hertzmann et. al's Image Analogies [Hertzmann et al. 2001] and Curve Analogies [Hertzmann et al. 2002] but various modifications are made to work for dynamically generated motion sequences. This metric has two key components: overall situational similarity and generated path coherency. These searches are done in parallel and combined in each step. Both comparisons are done over a given movement-history neighborhood n which is forty samples (one second) in our implementation.

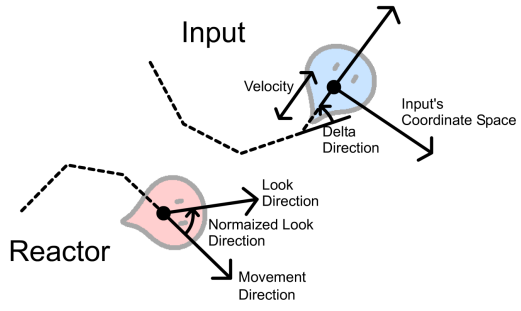


Figure 4: *Some of the data features calculated in our algorithm. All features except relative position are on both entities, but only shown on one for image clarity.*

4.3.1 Situation Similarity

The situation similarity is based on the relationship between the two entities. This step uses the relative position (input-character centric), relative look direction, and velocity features. It compares the n most recent pieces of user input and generated output from $< I, R >$ to a moving window of size n over the training data $< I^t, R^t >$ (Figure 5(a)). At each window location the features from each of the four paths form multi-dimensional vectors and corresponding vectors (I vs I^t, R vs R^t) are compared using Euclidean distance squared. These distances are then summed over the window, providing a measure of similarity at that window location. A smaller value represents a better match.

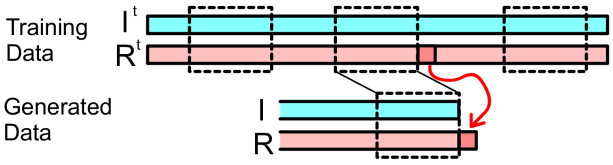
4.3.2 Generated Path Coherency

Done in parallel to situation similarity, this emphasizes the shape, style and features of the generated path R in relationship to the trained R^t while putting less emphasis on the relationship between entities I and R. This relationship is still important, however, as some aspects of coherency depend on the relationship between the entities, such as when the reacting entity wants to finish a circle around the main entity it must properly follow as the main character moves. This metric uses normalized look direction, delta direction, relative position, and velocity. When there is no training data that matches well to the current inter-entity situation (i.e., situation similarity is weak) generated path coherency helps to ensure a generation that matches the characteristics of R^t.

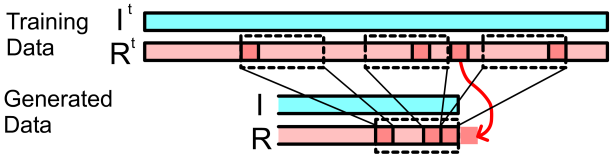
This metric compares the recently generated data from R over n to the regions in R^t that were used to generate the recent R (the entire R^t is not searched). That is, given recently generated elements R_k over the n, the source neighborhood in R^t that was originally used to generate R_k is compared to the most recently generated R (Figure 5(b)). The intuition here is to continue a patch from a previous similarity match if the current similarity match is weak.

4.3.3 Similarity Balancing

The Image Analogies algorithm [Hertzmann et al. 2001] combines the two similarity metrics by statically weighting them with a coefficient k to add bias; the metric with the best weighted score is selected for that step. This did not work with our application and resulted in a problem we call coherence loops: when coherence match is used to generate output for several consecutive steps then the result of generation, by design, will be increasingly similar to the training data. The improving coherence match is eventually exclusively used, with situation similarity being ignored, and the reacting entity starts to loop through sections of R^t. This issue does not occur in Image Analogies and Curve Analogies because all data is given at the beginning, allowing the use of multi-resolution approaches. Multi-resolution is difficult in our system, however, as we are generating in real time and cannot look ahead in our input data. To fix this problem, we tried to mesh results from both metrics in several ways (e.g., average, trend-vs-detail, staggered sample rate), but did not get satisfactory results. Our solution is given below, but this is a rich area for future work.



(a) Situation Similarity



(b) Generated Path Coherency

Figure 5: *In Situation Similarity the system compares recent real-time data to the entire dataset. In Generated Path Coherency the system only examines the regions of the training data recently used in generation. In both instances, the training data with the best feature match is chosen to generate the next output.*

We change the previously-static weighting coefficient k to a dynamic value that follows a target situation-similarity-to-coherency match ratio t. Then, k is automatically and continuously tuned each generation step to bias the results to reach t, so that over time we keep a balanced use of both metrics. In our implementation, we use a 1:1 target ratio. A similar algorithm is used in texture synthesis systems to match the overall color histogram [Kopf et al. 2007]. Once the best-matching source region is calculated, the data from R^t immediately following this source region is passed to the generation system. One problem with this balancing approach is noise. Instability in the similarity metrics (jumping between regions) and switching rapidly between situation similarity and coherence can cause large, rapid variations in the source data passed to the generation system, resulting in distracting rapid character movements. We explain how we deal with this in the next section.

4.4 Output Generation

The generation system receives the piece of training data (called the target data) to be used from the `BestMatch` function and generates the next entity output. The naïve approach is to simply copy this data directly to the output as in texture synthesis. The problem with this is that many features depend on entity history as well as the other entity (relative position, etc.) and it is impossible to solve a movement that matches all features. Also, when the training data jumps between drastically different states over consecutive steps this approach does not provide a meshing mechanism to generate intermediate data. That these jumps happen suggests that transitions are missing from the training data, and the generator function must provide a good approximation of the target features while handling discontinuities, all the while maintaining the texture, personality, and character that was demonstrated to the system.

Our generation approach, a key technical contribution of this paper, is to decompose the motion into its low-frequency (intentional move to certain relational position) part and high-frequency (texture of the motion) part and treat them separately.

4.4.1 General Trajectory Generation

The system generates motion using the relative position, normalized look direction, and velocity features. Normalized look direction is copied directly to output, and a vector is constructed to move the entity from its current location to the target relative position. This vector is scaled to the target velocity (Figure 6). Although this makes the entity move toward the target relative position rather than be at that position, the velocity scaling in combination with the high generation rate helps to create very convincing results.

Here we deal with the noise resulting from the `BestMatch` instability by applying a simple linear smooth (average) over a history of three samples. The results of this are very convincing and re-

sult in a more stable, consistent generation. The problem, however, is that by removing the high-frequency noise we also remove the high-frequency data, such as the movement detail and texture. We implemented a fix for this described in the next section.

Another problem with this system is that, even with smoothing, normalized look direction is very noisy. This happens because of the nature of the normalized look direction itself: if a character keeps a static look direction in the world, but rapidly changes movement direction, the normalized look direction (based on movement direction) changes rapidly between drastically different values. In particular, an entity moving rapidly forward and then backward has data that alternates between 0 and π. Our solution to this is to limit rate of change of the actual world-coordinate look direction. This lowers the amount of noise in resulting look direction, but some jitter remains. This is an important problem for future work.

4.4.2 Detail Incorporation

To restore the high detail information that was removed from the generation by smoothing we do frequency analysis using Haar wavelets, extracting the high-frequency detail from the target and directly incorporating it into the output. We apply Haar decomposition on the motion direction feature as this captures path texture irrespective of velocity. A single application of the discrete Haar decomposition scales our path data to half resolution and stores the removed high-frequency detail separately. This gives a frequency cut at $f_s/2$ where f_s is the sampling rate: given k cumulative decompositions, this cut is at $f_s/2^k$. The resulting k high-frequency datasets (one per decomposition) can be re-composed to form a single high-frequency-only signal that we use in our generation. Our system uses four-level Haar decompositions, a frequency cut of $f_s/16$, or about 2.5 samples per second. We found this to capture sufficient detail without affecting general trajectory.

The high-frequency data from the target is used to perturb the generated (but smoothed) trajectory. While smoothing compensated for `BestMatch` instability in trajectory generation, high-frequency source data cannot be smoothed in the same fashion and so this instability remains, where interweaving detail from rapidly alternating training locations results in noisy output not coherent to the training. Our solution applies the detail in patches of sixteen samples: a target"s source patch is used in subsequent steps until the end when a new patch is selected. These patches are only 0.4 seconds long so the delay between changed behavior and matching path detail is minimal and the results are satisfactory.

5 Evaluating Puppet Master

Our evaluation consisted of two parts. The first part of the study, the *artist* study, asked participants to design new behaviors using our system. The second part of the study, the *end-user* study, asked participants to interact with pre-modeled behaviors. Our general goals were to identify weaknesses in our algorithm and interface, to construct an initial picture of what the users think about the system, and to determine how much (and what sorts) of characteristics, emotions, and personality traits are captured by our system.

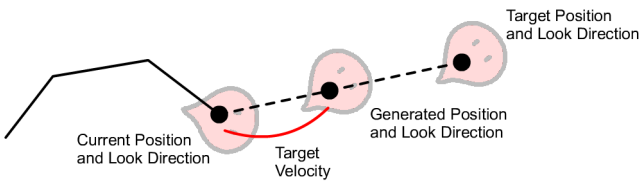


Figure 6: *Computation of general trajectory. To prevent drastic jumps, the reactor moves towards the target position with the given velocity taken from the training data.*

We initially conducted a pilot study to evaluate the study protocol and procedure. 5 participants (2 female, 3 male) joined the *artist* pilot and 2 participants (1 male, 1 female) joined the *end-user* pilot. These pilots exposed language and questionnaire wording that was confusing or strongly biased users toward particular responses.

5.1 Experimental Testbed

For the experiments we used the tabletop TUI interface (Section 3.2). It was run on a Pentium 4, 3.0 GHz PC, maintaining around forty frames-per-second for up to about 80 seconds of training data (a behavior generally requires less, as described below). For all experiments we used the graphics shown in Figure 7 which were static textures and were not animated other than changes in its location and orientation. We use a SMART Technologies 4'10" x 3'7" high-resolution (2800x2100) rear-projected tabletop display with plastic pucks on the top to control input (Figure 3). The pucks are tracked at 100fps by a six-camera Vicon motion-tracking system, and the character graphics are drawn on the tabletop surface directly below the pucks in real time.

5.2 Methodology

5.2.1 Participants

Twenty students (10 per study) from varying disciplines were selected from our university population and paid $15 for participation. All users reported some to extensive programming experience and strong confidence with computers. In the *artist* study (2 female, 8 male), four participants reported artistic experience with three having formal training and one identifying herself as an artist, and three users reported basic animation experience. Ages ranged from 19 to 32 (M=22.8, SD=3.8). In the end-user study (4 female, 6 male), nine participants reported artistic experience with five identifying themselves as artists, and four users reported animation experience (two extensive). Ages ranged from 19 to 27 (M=23.7, SD=2.71). All participants had no prior exposure to the system and no participants from the *artist* study took part in the *end-user* study.

5.2.2 Procedure

The purpose of the artist study was to see how general users can use our system to create interactive behaviors. Participants in this study, in roughly one-hour sessions, were first asked to design five particular interactive character behaviors given the following keywords: *lover, bully, playful friend, stalker,* and *afraid*. Participants completed a short written survey about the result and experience after each behavior. Following, we evaluated the internal validity of the design by loading the five created behaviors each participant created in a scrambled order (fixed across participants) and asking them to interact with, and recognize, each behavior. Participants

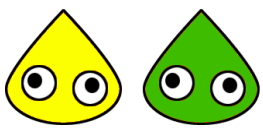


Figure 7: *The graphics used in our evaluation, designed generically to avoid conveying particular emotion or personality.*

were not notified ahead of time that they would be revisiting and evaluating their own designed behaviors.

The *end-user* study was conducted to observe how a general user reacts to the behaviors created using our system, and whether a sense of character emotion and personality will emerge. We subjectively selected five behaviors created by participants in the *artist* study (one per each of the five behavior types), and participants were asked to "interact with and explore the characters" for each behavior in a fixed order. Participants were asked to "describe the character" in a questionnaire. Particular care was given to avoiding affective or anthropomorphic language when presenting the task to the end users. Words such as "personality", "social", "behavior", "emotion", etc., were avoided. In the second part of the *end-user* study participants were asked to interact with a set of "other" behaviors which were in-fact a scrambled set of the previous behaviors. This time users were asked to match each of the behaviors to the list of "correct" behaviors as given in the *artist* study.

5.3 Artist-Study Results

Eight of ten users in the artist study properly identified and labeled all five of their own behaviors. Further, in 74% of the behavior cases (using 5 point Likert scale) users agreed or somewhat agreed that they were satisfied with the resulting behavior, and in 22% of the cases they neither agreed nor disagreed. The mean of the training time across users for accepted behaviors is 32.5 seconds (SD=18.0s, min=9s, max=85s). The average number of training trials (iterations) required before accepting a behavior was 1.7 (SD=0.9, mode=1 at freq.=56%, max=4 trials). The average amount of time a user spent testing a generated behavior before accepting it was 70.0 seconds (SD=68.2s). In 46% of the cases users disagreed that the generated behavior felt mechanical with 26% neither agreeing nor disagreeing. In 48% of the cases users agreed that the behavior felt human-controlled (42% only somewhat) with 26% neither agreeing nor disagreeing.

In the post-test questionnaire, on seven-point Likert scales, all 10 artist users agreed (5 strongly) that they enjoyed using the system, while 7 disagreed that the system was frustrating to use (1 strongly and 2 only somewhat), all users reported that the resulting characters were fun to play with (6 strongly and 2 somewhat) and 6 users reported that movement jitter was distracting. The only two users who failed to fully recognize their own designed behaviors were also the only two users who did not use puck orientation during behavior training, resulting in poor quality behaviors.

Four of the artist users were notably highly immersed in the design process. For example, some made exaggerated faces, noises, and spoke out loud to the characters while training. For example, an artist used themes from the movie Jaws while training the "afraid" behavior, and another commented "what a jerk!" when observing the designed "bully" character.

Users generally expressed excitement about and satisfaction with the capabilities of the system: "the system responded accurately and behavior was smooth, human-like, with a human touch", "it"s even a better stalker than I am!", "it almost looks as if someone is controlling it.", "it did exactly as I wanted! Very entertaining! (maybe it"s just me?)", "nailed it!", "I like it! I can see its bright future in entertainment, gaming, and teaching", "the playful friend is a hoot!" Several users also commented on the robustness of the system, and one user was excited that the system "even reacted consistently with what [he] thought of after the fact." Also, most users enjoyed the TUI tabletop system, with one user stating that it was "super easy and intuitive to operate. Instant results."

On the other hand, several users reported issues with the system, commenting on the resulting generation as well as the simplicity of our system: "it felt a bit mechanical with some movements", "as complexity of behavior rises it feels more mechanical", "if you pause to catch your breath, the system takes it as deliberate behavior", "I need to try more complicated behaviors", "this setup cannot interpret smaller actions that well", "he doesn"t have hands so I can"t punch", "difficult to imagine what one pretty slime does to bully another pretty slime." Further, six of the ten users had issues with occluding the Vicon markers on the controller puck.

5.4 End-User-Study Results

In the first part of the end-user study users were simply asked to interact and describe prototype characters (without being prodded to look at behaviors or emotions). Here, on a six point scale titled "the character felt..." ranging from "extremely mechanical" (1) to "somewhat mechanical" (3, 4) to "not mechanical at all" (6) the average response across all behaviors was 4.04 (SD=1.19,Mode=4 at 36% frequency). On another scale ranging from "a human is controlling it" (1) to "somewhat lifelike" (3, 4) to "not lifelike at all" (6), the average response was 3.4 with a mode of 5 at 24%.

To our pleasant surprise, out of the 50 cases (5 behaviors across 10 participants), characters were identified using the exact keywords used in the artist study 9 times, and 10 times using very similar words (for example, "girlfriend" instead of "lover", "naughty, trying to bug me" instead of "bully"). Out of the 10 participants, 2 did not match any behaviors, 2 matched 1 behavior, 3 matched 2 behaviors, 1 matched 3 behaviors, and 2 users matched 4 behaviors correctly. Furthermore, in the open-ended questionnaires 52% of all end-users behavior descriptions were using social and behavioral descriptions (28% purely social), 34% of all the descriptions were using mechanical language (18% purely mechanical), with 14% being roughly a half-half mix.

For the second part of the *end-user* study, participants were asked to match the five behaviors against the original keywords used. The results are given in Table 1, with the diagonal showing the number of end-users, out of 10, who managed to match the pre-designed behavior to its exact keyword.

Overall, at the final open-ended questionnaires, 4 users agreed that the characters actions were sometimes confusing (1 somewhat), 1 neither agreed nor disagreed, and 5 users disagreed (1 strongly, 1 somewhat). One strong observation throughout the entire *end-user* study is that users tended to see social characteristics and used anthropomorphic language. For example, end users mentioned that: "although they are just results of programming languages, they reflect social interactions and peoples" traits", "exhibit human behavior somewhat naturally at 2D level", "the guy who kept sucker-punching", "each one could bring to mind some real-life analogy", "he needs more confidence", "I liked the part when it came close

Matched to \ Actual Trained Behavior	Lover	Bully	Playful Friend	Stalker	Afraid
Lover	6	1	3	0	0
Bully	0	5	4	0	1
Playful Friend	4	3	2	1	0
Stalker	0	0	1	6	3
Afraid	0	1	0	3	6

Table 1: *How participants matched behaviors to original designs.*

to my character and kind of hugged it, kind of like a dog who is happy to see you", "He keeps trying to either hit you or kiss you", "like an annoying kid brother in my face", "he [the stalker] seemed like he wanted to approach me, but he was too shy", "facing it and watching it panic like it had been discovered somewhere where it shouldn"t be was fun", "she [playful friend] is like a little sister who wants to talk to me."

User were asked on the final questionnaire to describe the things they liked and disliked about each character. While some of these comments were analysis oriented, such as "it felt very mechanical because I could figure out what it was going to do next" or "actions were vague, subject to interpretation". Many of the comments referred to the participant"s opinion of the character"s personality. For example, for the afraid character (which stayed away from the participant"s character) one user wrote "I didn't really like anything, didn"t even give me a chance to get to know him". Similarly, participants mentioned behavioral-personality attributes when they were asked what they disliked about characters, for example "tries to invade my personal space. I like a nice personal space bubble", or "it doesn't feel friendly!"

Similar to the artist study, some participants in the end-user study also commented that the characters felt a bit fake when the jitter was too noticeable and several participants complained that the personalities were too simple and wanted additions: "the personalities were very blunt, they were easy to see", "I wish they could touch each other".

All end-user participants reported enjoying the experiment (6 strongly agreeing). 7/10 users reported the pucks frustrating to use (all of these users commented on how easy it was to occlude the Vicon markers), with the remaining 30% disagreeing or strongly disagreeing. However, several users commented that the table was "easy to use" and "intuitive".

6 Discussion

We believe that these results help to support several of our claims about Puppet Master. The fact that 80% of our artist study participants recognized 100% of their own behaviors and reported strongly that they were satisfied with the results suggests that our algorithm successfully supports some level of artistic expression and captures a sufficient and valid set of details and personality-related characteristics for recognition by the designer. Also, this combined with the fact that this was accomplished with no training at on average 32.5 seconds shows that, for even the slower outliers (such as the 85-second case), our algorithm allows artists to create recognizable behaviors in significantly less time than it would take to program it. Finally, that this was accomplished in on-average 1.7 training attempts for the first time using this system shows that the designers were able to satisfactorily create their behaviors fairly easily and without several iterations, but also that the real-time retraining and generation enabled the artists to flexibly explore design possibilities. For the two users who did not use orientation in their training and got weak results, this exposes a drawback related to the orientation input in our algorithm. We believe that a short orientation-input training session could have resolved this issue, or alternatively that the system could be modified to focus more on orientation-independent features.

The end-user part of our study demonstrated that without telling users to look for personalities, in 38% of the cases not only that behaviors emerged, but they closely matched the artist keywords based on motion only (Table 1). We believe that this supports our claim that our algorithm captures the personality and style of the demonstrated behavior. Further, the results in Figure 3 seem to hint at crosstalk between similar behaviors: for example, afraid and stalker are often mistook for each other while lover, bully, and friend are rarely mistaken for stalker or afraid. This shows that, even in the cases where behaviors are not matched properly, there is still a strong component of feeling and style captured from the demonstrated data.

Finally, both studies suggest a strong sense of user engagement. The combination of the explicitly positive study results, the verbal excitement, as well as the fact that people tended to extensively use social and anthropomorphic language suggests that the participants were interested and mentally involved with the design process and the characters they interacted with.

7 Limitations and Future Work

Our current implementation does not handle behavioral dynamics over a larger time scale. For example, our current algorithm will fail to accurately represent an angry character gradually calming down. We believe that such challenges can be approached by exploring high-level features and multi-resolution similarity searches that can potentially represent higher-level behavior patterns. It is also useful to consider how this work can combine with other behavior models and systems for a multi-level solution, and to understand the absolute limitation of our approach, that is, when high-level behavior changes may be better designed using scripting and explicit states.

Extending our system to account for environmental issues such as barriers (wall, tree, etc) or terrain type would dramatically improve its versatility and power. Related to this, extending our system to several simultaneous entities (not just two) is an exciting prospect that would allow swarm-like behavior with individual personalities through demonstration, for example, to train a group of archers and knights to storm a castle. There is a system that learns crowd behavior from examples [Lerner et al. 2007], but they mainly focuses on collision avoidance. We are interested in allowing artists to interactively design more intentional crowd behaviors.

An important future work is to apply our method to the design of real robot behaviors. As an initial step, we are thinking of replacing physical pucks and tabletop displays with mobile robots such as curlybot [Frei et al. 2000]. The physical movement of these robots can give stronger impression of personality and emotion than virtual characters. Physical motion introduces severe constraints on motion synthesis (e.g. robot cannot jump to distant position), but many of the techniques developed in this work such as similarity-coherence balancing and trajectory-texture separation should be applicable to real robots, too.

Our ultimate goal is to design all aspects of character behavior, not just locomotion, responding to various input such as dancing to music, sword-swing against an opponent, and meow noise of a cat responding other meow noise. These are all exciting prospects for extending this work.

8 Conclusion

Believable, convincing, and stylistic interactive behavior is an important aspect of any computerized entity that must interact with humans, such as avatars, video game characters, or even robots. Traditionally, the creation of such a system has been left to logical, step-by-step computer algorithms; tools generally out-of-reach for non-technical artists and ill-designed for the creation of stylistic behaviors. In this paper we presented the first system that enables the programming of interactive behaviors by demonstration with real-time generation, making the creation of believable, stylistic interactive characters accessible to the non-technical artist.

Acknowledgements

References

BATES, J. 1994. The role of emotion in believable agents. *Comm. ACM 37*, 7 (July), 122–125.

BLUMBERG, B. M., AND GALYEAN, T. A. 1995. Multi-level direction of autonomous creatures for real-time virtual environments. In *Proc. SIGGRAPH '95*, 47–54.

BREAZEAL, C. 2002. *Designing Sociable Robots*. MIT Press.

CYPHER, A. 1991. Eager: programming repetitive tasks by example. In *Proc. CHI '91*, ACM, NY, USA, 33–39.

DONTCHEVA, M., YNGVE, G., AND POPOVIĆ, Z. 2003. Layered acting for character animation. *ACM Trans. Graph. 22*, 3, 409–416.

FJELD, M., BICHSEL, M., RAUTERBERG, M., AND I PRESS, 1986. Build-it: a brick-based tool for direct interaction.

FREI, P., SU, V., MIKHAK, B., AND ISHII, H. 2000. curlybot: designing a new class of computational toys. In *Proc. CHI '00*, ACM, NY, USA, 129–136.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *Proc. SIGGRAPH '01*, 327–340.

HERTZMANN, A., OLIVER, N., CURLESS, B., AND SEITZ, S. M. 2002. Curve analogies. In *Proc. EGRW '02*, 233–246.

IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. 2005. Spatial keyframing for performance-driven animation. In *Proc. SIGGRAPH '05*, 107–116.

IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. F. 2005. As-rigid-as-possible shape manipulation. *ACM Trans. Graph. 24*, 3, 1134–1141.

ISHII, H., AND ULLMER, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proc. CHI '97*, ACM, New York, NY, USA, 234–241.

KOPF, J., FU, C.-W., COHEN-OR, D., DEUSSEN, O., LISCHINSKI, D., AND WONG, T.-T. 2007. Solid texture synthesis from 2d exemplars. *ACM Tran. Graph. 26*, 3, 2:1–2:9.

LEE, J., AND LEE, K. H. 2004. Precomputing avatar behavior from human motion data. In *Proc. SCA '04*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 79–87.

LERNER, A., CHRYSANTHOU, Y., AND LISCHINSKI, D. 2007. Crowds by example. *Comp. Graphics Forum 26*, 3, 655–664.

MAES, P. 1995. Artificial life meets entertainment: lifelike autonomous agents. *Comm. ACM 38*, 11, 108–114.

MATSUI, D., MINATO, T., MACDORMAN, K. F., AND ISHIGURO, H. 2005. Generating Natural Motion in an Android by Mapping Human Motion. In *Proc. IROS '05*, IEEE, 1089–1096.

MAULSBY, D. L., WITTEN, I. H., AND KITTLITZ, K. A. 1989. Metamouse: specifying graphical procedures by example. In *Proc. SIGGRAPH '89*, ACM, NY, USA, 127–136.

NORMAN, D. A. 2004. *Emotional Design*. Basic Books, NY.

RAFFLE, H. S., PARKES, A. J., AND ISHII, H. 2004. Topobo: a constructive assembly system with kinetic memory. In *Proc. CHI '04*, ACM, NY, USA, 647–654.

REEVES, B., AND NASS, C. 1996. *The Media Equation*. CSLI Publ., UK.

REYNOLDS, C. W. 1987. Flocks, herds and schools: A distributed behavioral model. In *Proc. SIGGRAPH '87*, ACM, NY, 25–34.

SHARLIN, E., WATSON, B., KITAMURA, Y., KISHINO, F., AND ITOH, Y. 2004. On tangible user interfaces, humans and spatiality. *Personal Ubiquitous Comput. 8*, 5, 338–346.

THORNE, M., BURKE, D., AND VAN DE PANNE, M. 2004. Motion doodles: an interface for sketching character motion. In *Proc. SIGGRAPH '04*, ACM, New York, NY, USA, 424–431.

WILEY, D. J., AND HAHN, J. K. 1997. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications 17*, 6 (/), 39–45.

WOLBER, D. 1997. Pavlov: an interface builder for designing animated interfaces. *ACM TOCHI 4*, 4, 347–386.