Published in the Proceedings of Computer Graphics International 2002, 1-5 July, Bradford, UK

Experimental Investigation of Linguistic and Parametric Descriptions of Human Motion for Animation

Jason Harrison   Kellogg S. Booth   Brian D. Fisher
Department of Computer Science
and
Media and Graphics Interdisciplinary Centre
University of British Columbia, Vancouver, BC Canada

Abstract

Computers process and store human movement in a different manner from how humans perceive and observe human movement. We describe an investigation of the mapping between the linguistic descriptions people ascribe to animated motions and the parameters utilized to produce the animations. The mapping is validated by comparing both the linguistic and parametric descriptions to similarity measures obtained during experiments in which subjects made judgements about the similarity of pairs of animated motions. Analysis of the experimental data revealed that similarity judgements are not true distances (they do not form a metric space), but statistical tests and principal component analyses of the linguistic and parametric descriptions indicate correlations with the similarity judgements that provide the basis for mappings between the two descriptions. Although there are significant individual differences in the mappings across subjects, there is some indication that our methodology can be extended to provide a more robust direct mapping from linguistic descriptions to at least initial approximations of the corresponding parametric descriptions of the motion for use in computer animation systems.

Keywords: Computer animation, human figure animation, human movement, judgement of human movement, description of human movement, movement perception.

1. Introduction and Overview

The leading paradigm used by computer animation systems employs three tightly coupled models. These models, used in movies such as *Toy Story*, *Final Fantasy: The Spirits Within*, and *Monsters Inc.*, are (1) a set of time signals $\mathscr{Q}(t)$ that specify the kinematics of the movement, (2) a mapping $\mathscr{A}$ between $\mathscr{Q}(t)$ and the position, orientation, and posture of the human figure, and (3) a “costume” or “visual appearance” that specifies the outer appearance of the human body.

In sharp contrast to the exactness of computers, it is not well understood how we visually perceive human movements. It is believed that we utilize the motor control centers of our brains to recognize and interpret the movements of others. However, we do not know how observed movements are encoded or how they are translated into linguistic descriptions. Neither do we understand the process we use to translate "mental images" of movements into physical movements.

In order to build higher-level computer animation tools for selecting, specifying, or modifying movements represented by computer models we need to know how the parameters of a movement, $\mathscr{P}(\mathscr{Q})$, affect our perceptions and judgements. We present results from two participant-based experiments that gathered information on the relationships between three motion spaces: the first motion space is the "mechanical motion space," a vector space of motion signals, $\mathscr{Q}(t)$; the second motion space is the "psychological motion space" in which humans encode and organize motions according to their features; and the third motion space is the "linguistic motion space" that humans use to describe movements using words.

In this paper we describe our initial findings about what we believe will eventually be a novel approach for specifying human motion in computer animation systems. Unlike traditional techniques that provide animators with tools that deal directly with the mechanical motion space in which motions are represented parametrically, our approach tries to bridge between the mechanical motion space and the more intuitive linguistic motion space in which motion is described using adjectives and adverbs. The bridging is accomplished using a hypothesized intermediate psychological motion space, based on human perception of motion and human judgements of motion similarity.

After providing a conceptual framework for relating the three motion spaces, we provide a brief summary of previous work including earlier attempts to provide animators with linguistically based tools for specifying motion. This is followed by a description of two experiments that were conducted to determine the nature of the mappings between the three motion spaces. The results of the experiments, and how those results provide a basis for future work that will refine our technique to produce practical tools for animators are then discussed.

The contribution of this work to the field of computer animation is three-fold:

(a) The conceptual framework relating the three motion spaces, as validated by our experiments, is a sound basis for mapping the more intuitive linguistic descriptions of motion to the commonly used parametric descriptions employed in many animation systems.

(b) The experimental methodology we have developed can be used to extend our preliminary findings to a more robust set of mappings covering a wider range of motions.

(c) The need to account for individual differences when calibrating the mappings suggests a number of research questions that should be pursued.

The conceptual framework and experiments reported in this paper were part of the research described in the first author's dissertation [9]. In this paper we describe the components of the conceptual framework, two experiments that were conducted to validate the framework, and the main conclusions that were drawn from the research.

2. Three Motion Spaces for Describing Human Movement

Many computer animation researchers focus on the technical aspects of computer animation systems such as increasing the realism of the visual models, refining techniques to record the movement of human actors, and building algorithms to assist in the editing of recorded movements. In other words, research has tended to focus on the limitations of computer based representations of human movement rather than on higher-level techniques for specifying and adjusting motions. For example, if an animator wanted to adjust the style of a movement while not affecting the gross path of the movement — making a movement more "happy," "awkward," "drunk," "graceful," or "fast" — almost none of the prior work in computer animation is applicable.

Our ultimate goal is to provide a direct and automatic mapping between everyday linguistic descriptions of human motion ("a sad, slouching man walking despondently down the road") to the mechanical (kinematic) description required by the underlying modeling and rendering software. This is beyond our current capability, but the work described here is a first step because it provides a framework for defining the mapping between the linguistic and mechanical motion spaces.

In order to build higher-level computer animation tools for selecting, specifying, or modifying movements represented by computer models we need to know how the parameters of a movement, $\mathscr{P}(\mathscr{Q})$, affect our perceptions and judgements. This requires knowledge of computer animation, human-computer interaction, and visual psychophysics. Thus, it is useful to introduce three different types of motion spaces to assist in our discussion of the relationship between the parameters of movement and the perceptions and judgements formed by a human observer.

The first motion space is the standard "mechanical motion space," a vector space of motion signals $\mathscr{Q}(t)$ that describes the kinematics of movement. Computer animation tools operate in this space. It remains an open problem as how to define a basis for $\mathscr{Q}(t)$ which would allow us to interpolate two motion signals while maintaining constraints such as foot contacts. Given this problem, we use the set of input parameters $\mathscr{P}(\mathscr{Q})$ to a kinematic walk generator that are used to define the motion signal $\mathscr{Q}(t)$. For example, if $\mathscr{Q}(t)$ is the motion signals for a walking motion, $\mathscr{P}(\mathscr{Q})$ includes joint angle limits, walking speed, step frequency, stride length, etc.

The second motion space is a hypothesized space in which humans encode and organize motions. We call this space the "psychological motion space." Although we know little about the structure or properties of this space, we can hypothesize that judging the similarity of two movements requires the computation of a "distance" between them. Although we consider this space to be pre-categorical, when we (as humans) categorize a movement as belonging to the class "running," "walking," or "throwing," we probably use the shortest distance between the movement and exemplars of these different classes [19].

The third motion space is also a conceptual space in which humans describe movements using words. This "linguistic motion space" contains attributes in which concepts such as "slower" and "bouncy" are defined. This is also the space used to interpret the labels on the user interfaces of computer animation tools, although there is little reason to believe that there is a solid basis for choosing the labels that are

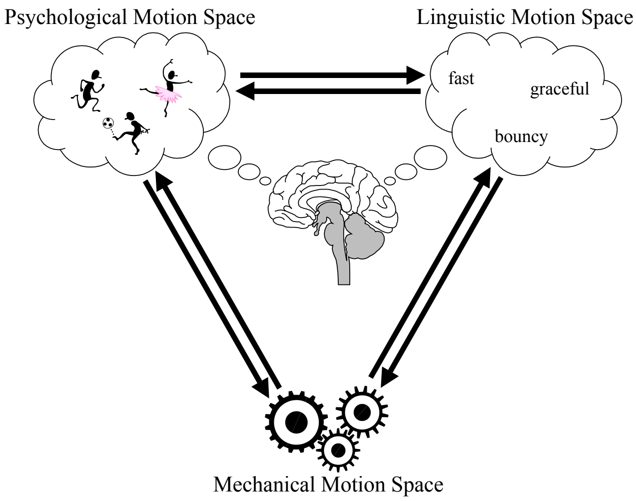

Figure 1: The three motion spaces used to discuss the relationship between the parameters of a movement and the perceptions and judgements formed by a human observer. Computer animation tools usually operate in the mechanical motion space, whereas humans perceive and code the features of movements in the psychological motion space and they describe movements with words in the linguistic motion space.

currently in use.

Figure 1 illustrates the relationships between these three motion spaces. One contribution of our work is the investigation of the form of the mappings between these three space. The mechanical motion space is a vector space, which makes it particularly amenable to computation — one of the main reasons it is used as the basis for most animation systems. The other two spaces are not vector spaces, but may have some structure that can be exploited. If they were vector spaces, the problem would be easy. We would simply choose convenient bases in each space and use those to define the mappings between them.

The linguistic space is defined by a number of categorical dimensions such as fast-slow, flexible-stiff, smooth-bouncy, etc. But these almost certainly do not form a vector space in the mathematical sense of the term. Even more problematic is the psychological space in which similarity-dissimilarity judgements are made. Ideally dissimilarity would have the properties of a distance measure, but it is not likely that human judgements form a metric space, much less a vector space. So we are left with the problem of determining the extent to which we can map between these three ways of describing motion in a manner that is useful for computer animation.

Computer animation programs represent human movement using a three-component model: time signals $\mathscr{Q}(t)$ specifying the movement of the figure, a mapping ($\mathscr{A}$)

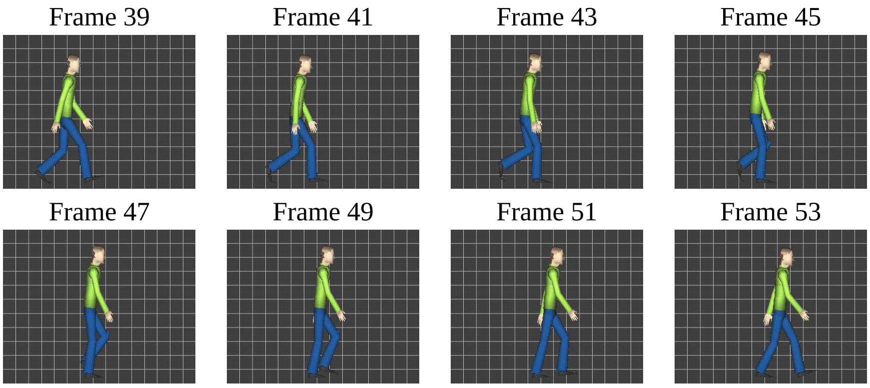

Figure 2: Sample frames from a computer animation display of a human walking movement.

between $\mathscr{Q}(t)$ and the position, orientation and posture of the human figure, and the visual appearance of the human figure (skin, clothing, and other surface attributes). This low-level representation is suitable for highly trained animators who are experienced at translating high-level natural language descriptions of motions (*e.g.*, scripts and storyboards) into manipulations of motion signals, but it is not very intuitive for non-animators.

Motion signals are transformed using the mapping $\mathscr{A}$ and the visual appearance of a human figure into a visual presentation termed a *computer animation display*. Computer animation displays of human figure animations rapidly create and present images ("frames") that show successive poses, positions and orientations of a human movement. Computer animation displays allow changes in viewpoint and arbitrary playback rates. Figure 2 shows eight frames from a computer animation display of a human walking movement.

As noted before, our ultimate goal is to allow animators to directly specify motion using the linguistic motion space, rather than requiring them to perform the tedious mapping to the less intuitive parametric description of the mechanical motion space. This is ambitious. It is likely that this will never be achieved in its entirety, but even a system that gets "most of it right" would substantially improve the process of computer animation. If skilled animators could get "first cut" motion sequences that were almost right, they could concentrate their efforts on the final tailoring of the details. This is a more achievable goal, and one that we believe we have made some progress on.

In the next section we survey earlier work on developing higher-level techniques for describing human motion for computer animation, emphasizing recent attempts to specify motion using linguistic or emotional descriptions.

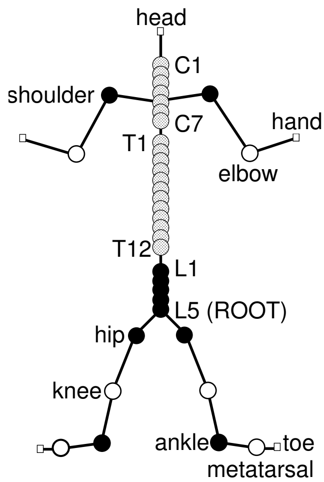

Figure 3: An articulation approximating the skeletal structure of the human body. Open (white) circles indicate hinge joints with only one rotational axis. Shaded (gray) circles indicate joints with two rotational axes. Dark (black) circles indicate joints with three rotational axes.

3. Related Work

In a computer program, the mapping $\mathscr{A}$ typically involves the specification of the jointed skeletal structure of a human. Originally suggested by Burtnyk and Wein [3], this structure is usually represented by an acyclic hierarchical articulation with the root at the hips, the approximate center of mass and movement. Zeltzer [23] presented a method for defining the articulation using a compact notation and Chadwick *et al.* [4] discussed the attachment of the visual form including deformable muscles, fat, skin, and clothing.

In computer animation applications, the human body requires from seventeen to over one hundred joints as illustrated in Figure 3. Some joints can be modeled as simple hinges with one rotational axis while other joints are modeled as "ball and socket" connections with three rotational axes.[1]

The movement of the human figure — that is the change in posture, position, and orientation of the articulation — is specified by the elements of $\mathscr{Q}(t)$, one function for each rotational or simple translational joint axis, and six functions specifying the position and orientation of the articulation as a whole. For classes of motions, such as cyclical symmetric straight-line walking motions, we can summarize the motion signals of a specific motion from the class with a set of motion parameters, $\mathscr{P}(\mathscr{Q})$. For walking motions, $\mathscr{P}(\mathscr{Q})$ includes parameters such as walking speed, stride frequency,

[1] Many of the joints are in the hands. Excluded from our definition of $\mathscr{A}$ is the ability to include facial expressions in $\mathscr{Q}(t)$ since these are not necessary for gross human body motions.

stride length, and joint angle limits of the arms, spine, hips, legs and feet.

For our work we used Bruderlin's `Walker` and his "Akira" visual form for the mechanical motion model [2]. Scaled values of `Walker`'s twenty-nine parameters were used to define $\mathscr{P}(\mathscr{Q})$. These parameters are: walk velocity, step length, step frequency, shoulder rotation, degrees of arm swing, angle arms are raised from the sides, elbow rotation (min, max), torso tilt (forward/backward), torso sway (side to side), torso rotation (left to right), lateral displacement, rotation of the pelvis about the vertical axis and list from side to side, bounciness on foot strike as the supporting knee bends, amount of over stride, hip flexion as leg swings forward, knee angle during swing, midstride and impact, stride width (distance left to right of foot placements), angle of foot with respect to straight forward ("pigeon or duck toes"), and whether the heel or toe strikes the ground first, as shown in Figure 4.

Bruderlin's `Walker` is one attempt to allow the specification of human movements at a higher level than a full specification of the motion signals, *i.e.*, a particular parameter set $\mathscr{P}(\mathscr{Q})$ for $\mathscr{Q}(t)$. Manipulations of `Walker`'s slider-based user interface produces walks with different styles that are always dynamically stable and physically plausible. Subtle constraints between the parameters are automatically adjusted by `Walker`, *e.g.*, increasing the walking speed increases the stride frequency and stride length, and leans the figure forward. Nevertheless, Bruderlin's description is still in a mechanical motion space.

We are interested in creating techniques that work at the level of movement quality and style. Other researchers have presented techniques for editing the quality of an existing motion or blending two motions with different styles to create intermediate styles. For example, Loomis *et al.* recorded the movements of native American Sign Language signers and then created synthetic signs and manipulated their movement quality independently of spatial path [13].

Much later, Unuma, Anjyo, and Takeuchi attempted to use Fourier analysis to represent cyclical motions such as running and walking [22, 21]. By interpolating and extrapolating between the Fourier coefficients they were able to create motions containing "emotive" content such as happy or sad walks, or normal and brisk walks. Their goal was to be able to extract the "briskness" factor from a walk and add it to a run to create a "brisk run." Along a similar line, Amaya *et al.* applied signal processing techniques to create an "emotional transform" capable of determining the difference between a neutral and emotional movement as contained in the timing and amplitude differences of the motions [1].

Rose and Cohen presented a system that parameterizes cyclical motion captured "verbs," such as walk, with "adverbs" that alter the motion, enabling them to create a continuous walking motion up and down hills, and turning left and right [20]. They could also add emotive variations such as happiness and sadness. The target application for their work was the real-time control of a digital puppet or video games character, with smooth transitions between motion verbs and adverbs.

Perlin presented a method to generate motion with "personality" by using expressions containing pseudo-random noise functions to generated the joint angle motions [16]. Also aimed at real-time motion generation, Perlin's system allows the operator to control a character while specifying particular moods and attitudes to be

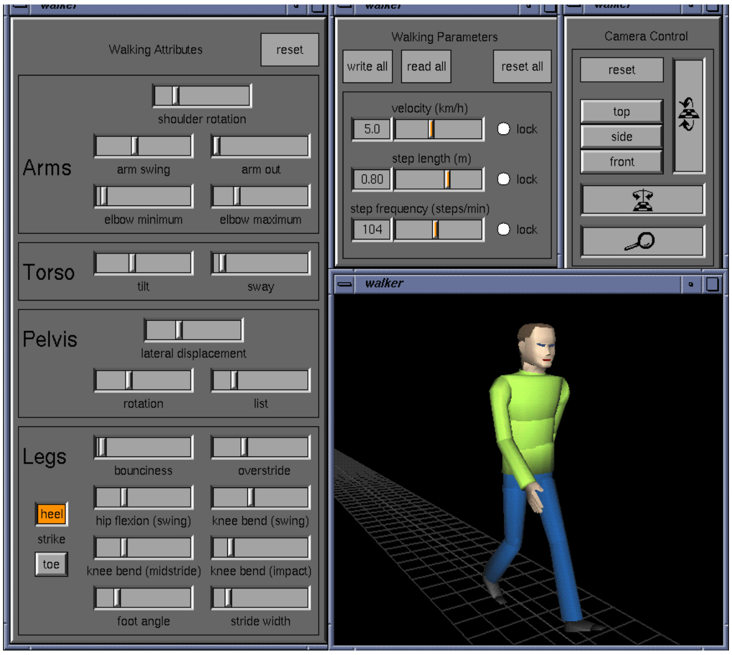

Figure 4: Screen shot of Bruderlin's `Walker` showing the output window and controls. The program uses twenty-nine inputs to vary the style of the walk by computing the motion of thirty-six joints totaling eighty-three degrees of freedom. The inputs are: walk velocity, step length, step frequency, shoulder rotation, degrees of arm swing, angle arms are raised from the sides, elbow rotation (min, max), torso tilt (forward/backward), torso sway (side to side), torso rotation (left to right), lateral displacement, rotation of the pelvis about the vertical axis and list from side to side, bounciness on foot strike as the supporting knee bends, amount of over stride, hip flexion as leg swings forward, knee angle during swing, midstride and impact, stride width (distance left to right of foot placements), angle of foot with respect to straight forward ("pigeon or duck toes"), and whether the heel or toe strikes the ground first.

conveyed. Kinematic constraints prevent “impossible” transitions.

All of these approaches use kinematics, as represented by the motion signals, rather than dynamics to model the motions. In contrast, Phillips and Badler presented a method of maintaining dynamic constraints such as balance and stability, which they consider “characteristics of human-like movement,” while the animator specified goal-oriented motions [17]. Popovic and Witkin also employed dynamics rather than kinematics, but generally focused on manipulating the spatial path of a motion while maintaining it’s existing movement qualities [18].

The weakness of these and similar techniques is that they are difficult to evaluate and there have been few attempts to validate the effects of the proposed manipulation methods. We are interested in descriptions of motion that are directly related to human perception of motion, and not tied to the underlying mathematics employed by the computational algorithms. One possible evaluation technique is to employ Certified Movement Analysts (CMAs) to estimate factors such as the Effort components of the resulting movements. CMAs are trained and certified by the Dance Notation Bureau [7], and employ a formal descriptive methodology developed by Rudolf von Laban [11, 12]. Laban’s four effort elements[2] attempt to capture the interactions between the spatial path and temporal elements of a movement that often occur in coordinated movements of the whole body.

One of the few computer animation systems evaluated by CMAs is Chi’s PhD Thesis on creating motion specified through a combination of keyframing and Laban’s Effort descriptors [5, 6]. Chi created several video sequences of a character moving its arms, and then presented these sequences with and without effort components. Four CMAs viewed the video sequences twice, the first time to familiarize themselves with the character and its movement, and a second time to judge which of the four Effort elements were present in each segment to what extent (positive, not present, negative). The CMAs judged at least 53% of the Effort elements correctly — that is, in agreement with the input settings of Chi’s system, with the in-house consultant judging 76.6% correctly. The majority of the incorrect judgements were “not present” judgements rather than opposite to the intended effort element.

Although Chi achieved moderate success with her system for interpolating keyframed movements, attempting to use Laban’s effort system to translate between computer and human representations of motion is very difficult if not impossible. Because Laban’s effort elements are a notational system used to communicate stylistic patterns of whole body movements between two humans they require experience of how humans move and code movement. While we can approximate the movement of humans by using $\mathscr{Q}(t)$ and $\mathscr{A}$ we do not yet know how humans code movements mentally or how much they rely on their own experiences of generating movement to code the observed movements of others.

Another technique is to have participants in an experimental setting categorize the resulting movements according to their emotional content. This technique has been used by Paterson and Pollick to determine the role of velocity in affect discrimination [15, 14]. Using both movements created by actors attempting to express different

[2]The elements are indirect versus direct, light versus strong, sustained versus quick, and free versus bound.

affects and neutral movements manipulated with signal processing similar to Amaya *et al.*'s techniques [1], Paterson *et al.* had participants name the resulting affects associated with each movement. Their results indicate a strong effect of velocity on the modulation of affect.

A final possible technique would be to obtain direct comparisons between two movements, one created by a computer and one recorded from natural human motion, and use these as a basis for mapping the linguistic descriptions of the natural human movement to the parametric description of the closest corresponding computer-generated movement. We have taken an approach that is related to this, but which adds a third motion space. Rather than assume that humans rely solely on either a categorical linguistic motion space — such as Laban's Effort model or Paterson and Pollick's emotional affects — we also assume the existence of a non-categorical non-linguistic psychological motion space in which observers code, compare, and analyse human movements.

The goal of our research is to eventually build higher-level computer animation systems by determining the structure and the bidirectional relationships between the three motion spaces. In the next section we describe the first steps in augmenting this basic framework with such mappings through the experimental collection and analysis of judgements made by human observers of computer-generated human motion.

4. Description of Experiments

We conducted two participant-based experiments to collect information about the relationships between the three motion spaces. The mechanical motion space is defined by `Walker`'s parameters, $\mathscr{P}(\mathscr{Q})$. The psychological and linguistic motion spaces are unique to each participant and are defined by their descriptions and comparison of human motions. Although we analysed the relationships between the motion spaces within-participants, we also looked for commonalities between participants.

The psychological motion space of each participant is defined by their judgements of the similarity of pairs of motions. We used a method similar to those used by earlier researchers. In both experiments comparison trials presented on a computer monitor two walking motions, one after the other. A participant then recorded their judgement of similarity on a continuous scale with the "Similar" end of the scale coded as zero, indicating "no distance" between the two motions, and the "Dissimilar" end of the scale coded as one, indicating a "large distance" between the two motions. Figure 5 illustrates a motion comparison trial. Each trial took about ten seconds: seven seconds for the presentation of the gaits and two to three seconds for the participant to make a judgement and click on the "Finish Trial" button.

After completing all of the trials involving comparing motions the participants described each motion. These descriptions define the linguistic motion space of each participant. After presentation of a single walking motion the participant recorded a description of the gait on eight continuous scales labeled with pairs of words with opposite meanings: fast-slow, flexible-stiff, smooth-bouncy, young-old, energetic-tired, light-heavy, graceful-spastic, and normal-strange. Ratings on each scale were coded in the interval $[0,1]$. Figure 6 illustrates a motion rating trial. Each trial took about

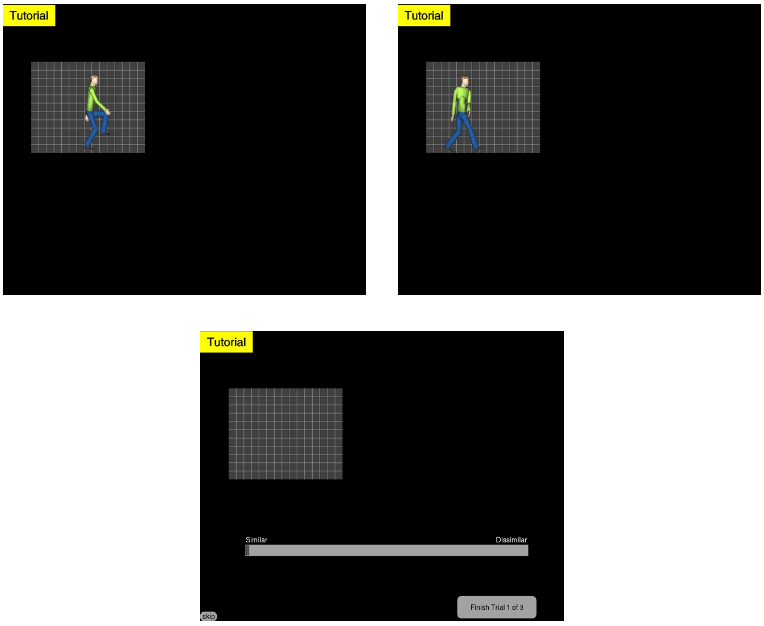

Figure 5: Three screen shots illustrating a motion comparison trial. *Top images*: The first gait is presented, then the second gait is presented. *Bottom image*: Next, the Similar-Dissimilar scale is displayed with the marker at the Similar end of the scale. After indicating a judgement of dissimilarity, the participant clicks on the "Finish Trial 1 of 3" button in the lower right hand corner of the screen. Participants were instructed to click on the "skip" button in the lower left hand corner of the screen only if they felt their attention drifted during the presentation of the stimulus and they failed to observe the walking motion. The "Tutorial" label in the upper left hand corner of the screen appears only during the training portion of the experiment.

fifteen seconds: four seconds for the presentation of the gait and about ten seconds for the participant to record ratings on each of the eight descriptive scales and click on the "Finish Trial" button.

All trials were replicated in random order in each of four blocks and participants were instructed to use the first block to learn the task of comparing or describing the gaits. Participants were shown all of the gaits used in an experiment as many times as they wished before beginning the first block of trials. Participants were naive to the hypotheses and were paid for their involvement.

Experiment One was a broad initial experiment performed to demonstrate the collection of similarity judgements and descriptions of the movements from human observers using a wide range of human walking movements. We systematically varied `Walker`'s motion parameters to create twenty-six gaits with as wide a variety as possible of walking motions with the constraint that all gaits had the same walking

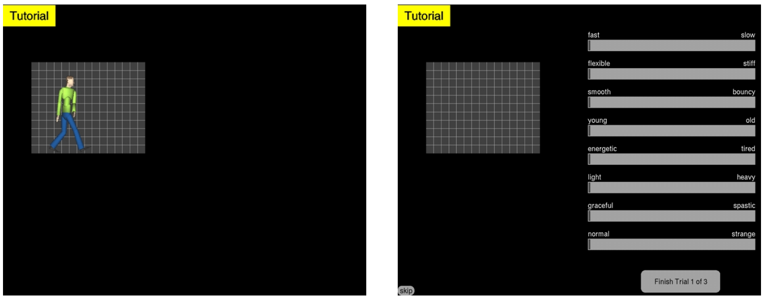

Figure 6: Two screen shots illustrating a motion rating trial. *Left to Right*: A gait is presented, then the eight rating scales were displayed. After indicating a description of the motion, the participant clicks on the "Finish Trial 1 of 3" button in the lower right hand corner of the screen. Participants were instructed to click on the "skip" button in the lower left hand corner of the screen only if they felt their attention drifted during the presentation of the stimulus and they failed to observe the walking motion. The "Tutorial" label in the upper left hand corner appears only during the training portion of the experiment.

velocity. Each gait was randomly paired with two other gaits to create fifty-two trials (repeated in each of four blocks). We recruited six participants: five were social dancers (three females and two males) chosen with the expectation that they would be able to apply their experience observing, imitating, and discussing human movements; the remaining participant was a non-dancer (male).

Experiment Two was an in depth experiment to determine the properties of the psychological motion space by using a narrower range of walking movements that included movements created by interpolating motion parameters. Figure 7 illustrates the two networks used to define the gaits. The first network is a triangle defined by three "primary" gaits, indicated with the filled circles. The gait parameters of the primary gaits were interpolated to create the gait in the center of the triangle. The second network is defined by two primary gaits on the ends of a line and the gait parameters of the primary gaits are interpolated to create three gaits along the line. The use of interpolation of gait parameters allows us to test the effect of interpolation in the mechanical motion space on proximities in the psychological motion space. Each gait was paired in both orders of presentation with all other gaits and itself within its network to create forty-one trials (sixteen plus twenty-five).

Reflecting our focus on the psychological motion space, we expanded the number and backgrounds of our participants to reflect the general population of individuals. We recruited thirty participants. They had experience in social dancing (seven females, one male), recreational running (nine females, eight males), or neither (one female, four males). We also introduced an additional experimentally controlled variable that specified the direction the figure walked across the screen: left to right — the same condition as in Experiment One (nineteen participants) — and right to left

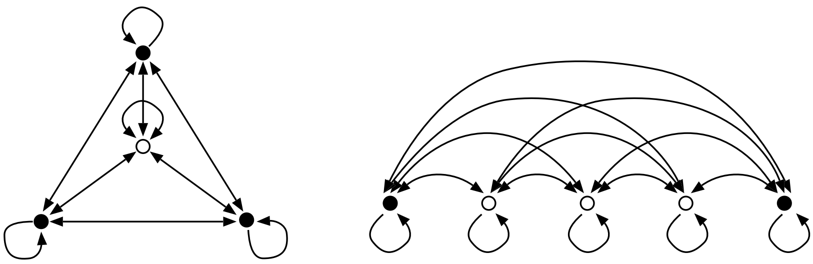

Figure 7: Networks of comparisons for primary gaits (filled circles) and interpolated gaits (unfilled circles). On the left, a triangle of three primary gaits with an interpolated gait in the center requires sixteen comparisons (arcs) to test the metric properties. On the right, two primary gaits are interpolated to create three new gaits, requiring ten comparisons to test metric properties and parameterization.

(eleven participants). We used a variety of measures, such as average judgement variance, badness of triangular fit, weirdness index (as computed using multidimensional scaling), and strength of correlation between similarity judgements and differences between ratings.

5. Analysis of the Properties of the Psychological Motion Space

The purpose of *Experiment Two* was to determine the properties of the psychological motion space — specifically by using networks of gaits which are compared in all possible combinations. By using interpolation of the motion parameters we sought to determine if the psychological motion space has metric properties. These properties are:

H_0^{1m} Non-degeneracy: only the self-distance is zero, and distances are never negative: $d(i,j) > d(i,i) = 0$.

H_0^{2m} Symmetry: distances between points are symmetric: $d(i,j) = d(j,i)$.

H_0^{3m} Triangular Inequality: sum of lengths of two sides of a triangle is never less than the length of the third side: $d(i,j) + d(j,k) \geq d(j,k)$.

If similarity judgements had the metric properties then we could continue to treat similarity judgements as approximations of the distance between motions. This would allow us to move beyond simple correlation between the spaces and begin to build models of how motion parameters are combined to form similarity judgements.

To test the first metric property we constructed trials in which each motion was compared to every motion within its network including itself. Our analysis of the motion comparison trials in which a motion was compared to itself revealed that the participants recorded a judgement more than one mark along the scale away from the similar end on average six times out of twenty-seven trials.[3] This indicates that $d(i,i) \neq 0$.

We also analysed the trials in which two different motions were presented and discovered an asymmetry in the pattern of judgements. We believe that if a participant recorded a similarity judgement for two different motions between the Similar end of the scale and the first mark on that scale that they thought that either they had seen the same motion twice or that they had seen two very similar motions. We term these judgements *mis-judgements*.

Participants tended to make a mis-judgement more frequently when the first motion was more "average" than the second. For example, when the center motion in the triangular network was presented first followed by one of the corner motions the number of mis-judgements was much higher — on the order of 50% — than the reversed order of presentation. These mis-judgements indicate that $d(i,j) = 0$ even when $i \neq j$.

We tested the symmetry of similarity judgements by testings if the average judgement $d(i,j)$ was equal to the average judgement in the reverse presentation order, $d(j,i)$ (within participants). The triangular network requires six of these comparisons, and the linear network requires ten. We used $\alpha = 0.05$ to indicate asymmetry of judgements. While each participant had at least one asymmetric average judgement, and the worst participant had four asymmetric average judgements, we did not feel that these asymmetries were strong enough to demonstrate violation of the symmetry property.

Finally, we tested the triangle inequality by counting the number of triangles formed using averaged similarity judgments that did not conform to the triangular inequality. We concluded that the triangular inequality did not hold because across the participants 14-30% of the triangles failed the triangular inequality.

We also compared groups of participants as defined by demographics (dancer/ runner/ neither, male/ female) and direction of walking according to the following measures: average judgement variance, badness of triangular fit, weirdness index (as computed using multidimensional scaling [8]), and strength of correlation between similarity judgements and differences between ratings. We found no large systematic differences between the populations of participants, indicating that perceptual similarity of motions relies on a basic perceptual process common to all humans.[4]

Since we interpolated the motion parameters to generate some of the gaits we examined the relationship between interpolation of the parameters and the resulting similarity judgements. While the relationship is not linear, we found it to be smooth and monotonic. This suggests that though the psychological motion space is not metric, it is locally well behaved.

To summarize, we tested similarity judgements against the metric properties and found evidence that the judgements do not conform to the non-degeneracy and triangle inequalities. This means that we cannot treat similarity judgements as approximations of the distance between gaits in a metric space. Comparison of the participants according to demographic and experimental variables indicated that comparing the motion of human walking figures is a process that does not require special training.

[3] There were eighteen regions along the Similar-Dissimilar scale demarcated with seventeen thin vertical lines. Thus the marks were placed 5.6% of the width of the scale apart.

[4] We could not test for interaction effects due to empty cells.

6. Analysis of Relationships Between Motion Spaces

In *Experiment One* the participants compared pairs of the gaits as to their similarity and described each of the gaits using eight descriptive rating scales. Combined with the parameters of the motions $\mathscr{P}(\mathscr{Q})$, we have three sets of relationships:

1. Mechanical and Linguistic Motion Spaces: which parameters $\mathscr{P}(\mathscr{Q})$ correlate with descriptions along each rating scale?

2. Mechanical and Psychological Motion Spaces: differences in which parameters correlate with similarity judgements of pairs of motions?

3. Psychological and Linguistic Motion Spaces: difference along which rating scales correlate with similarity judgements of pairs of motions?

Relationship between Mechanical and Linguistic Motion Spaces

For each participant, we determined which of the parameters influenced their ratings along the descriptive scales by computing the strength of linear correlations between the motion parameters and the descriptions. While the pattern of correlations differs for each participant, we found that each participant's ratings along each scale co-varied significantly with two to five parameters ($\alpha = 0.05$), indicating that each participant focused on particular body part movements to form their descriptions.

To determine which parameters most strongly influenced the ratings, we counted the number of participants with significant correlations for each combination of rating scale and parameter and list here those combinations which a majority of the participants (four or more out of six) had significant correlations:

fast-slow step length and knee swing (controls the amount the foot "kicks up" as it leaves the ground).

flexible-stiff upper torso rotation (vertical axis) and pelvis rotation (vertical axis).

smooth-bouncy bounciness (up and down motion of the body on each step).

young-old magnitude of arm swing (at shoulder), magnitude of elbow rotation, and knee swing.

energetic-tired magnitude of arm swing (at shoulder), magnitude of elbow rotation, tilt of torso, knee swing, average knee bend throughout stride.

light-heavy heel or toe strike on each step.

graceful-spastic magnitude of elbow rotation, torso sway (front and back tilt), bounciness, and hip swing (height of leg swing as foot swings forward).

normal-strange torso sway, bounciness, and hip swing.

These relationships are not too surprising — we carefully picked our set of descriptive scales and Bruderlin carefully picked `Walker`'s controls.

Relationship between Mechanical and Psychological Motion Spaces
We computed the strength of the correlations between the mechanical and psychological motion spaces by approximating the distance between pairs of gaits in the mechanical motion space by the difference in their parameters, as well as the differences of the Z-scores of the parameter computed using principal components analysis [10]. To summarize our findings, we found individual parameters that correlated stronger than 0.4 for all participants, and some correlations as strong as 0.6. The most popular parameters were those controlling the height the arms were raised from the sides of the body, the bounciness of the stride, and the fixed tilt of the torso. The strongest correlations between the similarity judgements and the principal components of the parameters were along the first principal component for all participants and ranged from 0.33 to 0.72 with a median value of 0.48. This indicates that the principal components of the parameters capture the interesting variation of the gaits in both the mechanical and psychological motion spaces.

We attempted to improve on the strength of these correlations by combining the principal components of the parameters by using the City-Block, Euclidean, and Dominance distance metrics (L^1, L^2, and L^∞, respectively). The strongest correlations varied from 0.41 to 0.73, with the Dominance norm (L^∞) producing the strongest correlations for four of the six participants.

Relationship between the Psychological and Linguistic Motion Spaces
For the relationships between the psychological and linguistic motion spaces we computed the strength of the correlations in a similar fashion. The strongest correlation between dissimilarities and differences along any single rating scale varied across the participants from 0.42 to 0.68 with a median of 0.58. The strongest correlations occurred along the young-old scale for four of the participants, and along the fast-slow and energetic-tired scale for the other two participants.

These correlations indicate that the rating scales can be used to indicate the similarity of gaits: similar gaits have similar descriptions and dissimilar gaits have dissimilar descriptions. As we saw before, the participants also tended to agree as to which parameters most strongly influenced each rating scale. This leads to our next question: do the participants' linguistic motion spaces have common features? To answer this question we computed the principal components of each participant's ratings. We normalized the ratings along each scale to have zero mean and unit variance. This gives each scale equal weight and tends to equalize the participants' interpretations of the scales.

Structure of the Linguistic Motion Space
We found that all of the participant's ratings could be compressed from eight rating scales to three or four principal components by taking only those components with a variance greater than one. This cut-off corresponds to a total variance accounted of more than 80%. Additionally we found several patterns across the participants' components:

- Five participants (#2, #3, #4, #5, and #6) had a very similar first principal component which is an average of all the rating scales and describes motions as either bouncy, spastic, strange, fast, heavy, energetic, young, stiff or as smooth, graceful, normal, slow, light, tired, old, flexible.

- Four participants had a single component dominated by the young-old, energetic-tired, flexible-stiff scales with the expected orientation: young, energetic, flexible versus old, tired, stiff. These components are the second principal components of #1, #3, and #4, and the third principal component of #6.
- The third and fourth principal components tend to be dominated by one or two scales — possibly reflecting the biases of each participant.

We computed the correlations between the similarity judgements and the Z-scores of the principal components of the descriptions and found for five of the participants the strongest correlation occurred along the first principal component (min = 0.47, max = 0.74, median = 0.62), and for the sixth participant the strongest correlation was 0.32 along her second component. We attempted to improve on the strength of these correlations by combining the principal components of the descriptions using the City-Block, Euclidean, and Dominance distance metrics. We were able to exceed the strength of the correlations for only half of the participants beyond the correlations computed along the components. For five of the participants the strongest correlations were produced using the City-Block norm, for the sixth participant the Euclidean norm produced the strongest correlations.

To summarize, in *Experiment One* for each participant we found moderately strong to strong linear correlations between the three motion spaces. We also found agreements between the participants as to which motion parameters influenced their similarity judgements and their descriptions, as well as common structures in their linguistic motion spaces. In each case, when we computed the principal components of the motion space we were able to find stronger correlations between the spaces using the principal components, rather than the original dimensions.

These results validate our experimental methodology and demonstrate that although we cannot assume that the participants would have similar biases as to the similarity of two gaits or how a gait should be described, the agreements between the participants suggest that higher-level human-computer interfaces to animation systems would not need to be custom tailored for each user, rather they could be “initialized” with a set of standard controls and then refined by adjusting the standard controls, or by choosing a set of controls that reflects the user’s perceptual biases.

7. Conclusions and Future Work

Computer graphics has benefited greatly from an understanding of human vision. For example, design guidelines based in knowledge of trichromacy, opponent colors, and JND measurements have informed the design of graphics hardware and software. For more complex stimuli and tasks, however, the basic perceptual abilities and bottlenecks are not well described. In the case of human figure animation, we know very little. This paper begins a process of exploring the motion spaces associated with human figure movement. We postulated two human-based representations of human figure movement, a linguistic space whose nature is determined in part by our ability to describe and categorize movements, and a psychological space more closely tied to perceptual events. Both contribute to our visual experience, and an understanding of both will support the creation of authoring and database management tools for character animation.

Experiment One examined the classification of gaits within the structure of pairs of opposite movement description terms. We found that the focus of attention varied among subjects, but that similar stimulus characteristics were salient in determining the classification of gaits, and that classification was somewhat consistent across most of our subjects and could be reduced to three to four principal components. *Experiment Two* explored the metric properties of the psychological space by asking subjects to make comparisons between a limited range of movements that were unlikely to span boundaries between multiple linguistic descriptors, but which were perceptually distinct. We concluded that similarity judgements do not have all of the metric properties. Our analysis was complicated by order effects of stimulus pair presentation, a not uncommon finding in complex perceptual tasks. Despite this, we did achieve some support for our hypothesis that the underlying perceptual evaluation was similar across subjects. The persistence of intersubject performance differences does suggest that future animation systems should be customizable not only for the user's preferences, but for their perceptual abilities and cognitive/linguistic structures as well. If our findings are correct, we can predict that this customization might be achieved by altering parameter values associated with the relative weights given to common perceptual cues without the need to add or substantially modify the nature of the cues themselves.

Our findings are preliminary, and more research will be needed before we begin to achieve our goal of creating animation environments capable of mapping between the intuitive linguistic descriptions of motions and computer representations by incorporating our understanding of the non-categorical aspects of human motions captured by the psychological motion space. Future experiments will explore the perceptual and linguistic spaces in more detail, and additional experimental controls for order effects should enable us to build models of the relationships between the three motion spaces with increasing predictive validity. In themselves, the present experiments have demonstrated that both the linguistic and psychological spaces have a level of structure and internal consistency that will support just such a parametric analysis.

Additional details of both experiments are available in the first author's doctoral dissertation [9].

References

[1] Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from motion. In Wayne A. Davis and Richard Bartels, editors, *Proceedings of Graphics Interface '96*, pages 222–229. Canadian Information Processing Society, Canadian Human-Computer Communications Society, May 1996. ISBN 0-9695338-5-3.

[2] Armin Bruderlin. *Procedural Motion Control Techniques for Interactive Animation of Human Figures*. PhD thesis, Simon Fraser University, School of Computing Science, March 1995.

[3] N. Burtnyk and M. Wein. Interactive skeleton techniques for enhancing motion dynamics in key frame animation. *Communications of the ACM*, 19(10):564–569, October 1976.

[4] John E. Chadwick, David R. Haumann, and Richard E. Parent. Layered construction for deformable animated characters. In *Computer Graphics (ACM SIGGRAPH 1989 Proceedings)*, pages 243–252, July 1989. Published as V23(4) of Computer Graphics.

[5] Diane Chi. *A Motion Control Scheme for Animating Expressive Arm Movement*. PhD thesis, University of Pennsylvania, Department of Computer & Information Science, June 1999. Also published at IRCS-99-06.

[6] Diane M. Chi, Norman I. Badler, and Janis Pforisch. Animating expressivity through effort elements. In *Siggraph'99 Conference Abstracts and Applications*, page 231, 1999.

[7] Dance notation bureau. http://www.dancenotation.org/, 2001. Phone: 212-564-0985, Fax: 212-904-1426.

[8] Mark L. Davison. *Multidimesional Scaling*. John Wiley & Sons, 1983.

[9] Jason Harrison. *Measuring and Comparing Human Walking Motions for Computer Animation*. PhD thesis, University of British Columbia, Department of Computer Science, December 2001. Available from `http://www.cs.ubc.ca/~harrison/Thesis/Doctoral/`.

[10] J. Edward Jackson. *A User's Guide to Principal Components*. John Wiley & Sons, Inc., 1991.

[11] Rudolf von Laban. *The language of movement; a guidebook to choreutics*. Boston, Plays, 1974.

[12] Rudolf von Laban and F. C. Lawrence. *Effort*. Macdonald and Evans, 1947.

[13] Jeffery Loomis, Howard Poizner, Ursula Bellugi, and John Hollerbach. Computer graphic modeling of american sign language. In *Computer Graphics (ACM SIGGRAPH 1983 Proceedings)*, pages 105–114, 1983. Published as V17(3) of Computer Graphics.

[14] Helena M. Paterson and Frank E. Pollick. Form and animacy in the perception of affect from biological motion. Poster at Vision Sciences Society First Annual Meeting, 2001. Available from the author: `http://staff.psy.gla.ac.uk/~helena/`.

[15] Helena M. Paterson, Frank E. Pollick, and Anthony J. Sanford. The role of velocity in affect discrimination. In *Proceedings of the Twenty-third Annual Conference of the Coginitive Science Society*, 2000.

[16] Ken Perlin. Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):5–15, March 1995.

[17] Cary B. Phillips and Norman I. Badler. Interactive behaviors for bipedal articulated figures. In Thomas W. Sederberg, editor, *Computer Graphics (ACM SIGGRAPH 1991 Proceedings)*, pages 359–362, July 1991. Published as V25(4) of Computer Graphics.

[18] Zoran Popović and Andrew Witkin. Physically based motion transformation. In Alyn Rockwood, editor, *Computer Graphics (ACM SIGGRAPH 1999 Proceedings)*, pages 11–20, Los Angelas, California, August 8-13, 1999. ACM SIGGRAPH.

[19] E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*. Hillsdale NJ: Lawrence Erlbaum, 1978.

[20] Charles Rose, Michael F. Cohen, and Bobby Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18(5):32–40, September/October 1998.

[21] Munetoshi Unuma, Ken Anjyo, and Ryozo Takeuchi. Fourier principles for emotion-based human figure animation. In Robert Cook, editor, *Computer Graphics (ACM SIGGRAPH 1995 Proceedings)*, Annual Conference Series, pages 91–96, Los Angeles, California, 6-11 August 1995, August 1995. ACM SIGGRAPH, Addison Wesley.

[22] Munetoshi Unuma and Ryozo Takeuchi. Generation of human motion with emotion. In Nadia Magnenat-Thalmann and Daniel Thalmann, editors, *Computer Animation '91*, pages 77–88, Geneva, Switzerland, 1991. Springer-Verlag.

[23] David Zeltzer. Representation of complex animated figures. In *Proceedings of Graphics Interface '82*, pages 205–211, 1982.