The Visual Computer manuscript No.
(will be inserted by the editor)

Animating with Style: Defining Expressive Semantics of Motion

Klaus Förger · Tapio Takala

pre-print version

Abstract Actions performed by a virtual character can be controlled with verbal commands such as 'walk five steps forward'. Similar control of the motion style, meaning how the actions are performed, is complicated by the ambiguity of describing individual motions with phrases such as 'aggressive walking'. In this paper, we present a method for controlling motion style with relative commands such as 'do the same, but more sadly'. Based on acted example motions, comparative annotations, and a set of calculated motion features, relative styles can be defined as vectors in the feature space. We present a new method for creating these style vectors by finding out which features are essential for a style to be perceived and eliminating those that show only incidental correlations with the style. We show with a user study that our feature selection procedure is more accurate than earlier methods for creating style vectors, and that the style definitions generalize across different actors and annotators. We also present a tool enabling interactive control of parametric motion synthesis by verbal commands. As the control method is independent from the generation of motion, it can be applied to virtually any parametric synthesis method.

Keywords Computer animation · Human motion · Motion style · Motion synthesis · Style vector · Feature extraction · Feature selection · Verbal description of motion style

Klaus Förger · Tapio Takala
Aalto University, Dept. of Computer Science
Otaniementie 17, 02150 Espoo, Finland
E-mail: klaus.forger@aalto.fi
E-mail: tapio.takala@aalto.fi

1 Introduction

An animator would often like to control virtual characters the way a theater director does, giving verbal commands rather than manipulating individual limbs like a puppeteer. Goal-oriented actions can be created with existing motion synthesis methods [7,12,15], even by scripting the requirements in natural language [14]. Different styles, meaning how actions are performed, can be produced with parametric and example-based methods [8,17,18,20,23]. However, controlling style with verbal attributes has received less attention. Many motion synthesis methods do not have a direct relationship between input parameters and the resulting styles. To fill this gap, we present a method that allows accurate control of motion style with high-level natural language commands. A similar approach has been applied in controlling color themes to create affective changes in images [24].

We define motion style to be a visually recognizable aspect of captured or synthesized motion. Furthermore, we define absolute style as one that can be perceived from individual motions and relative style as that perceived from differences between motions. Motion styles can be modeled numerically or described with natural language. In this work we seek correspondences between these two, in order to computationally define, identify and control styles in animation.

Judgements about styles are more vague and subjective than about goal-oriented actions. For example when a character tries to reach an object, we can measure if the hand touches the object, but it is less clear if the hand motion is seen as aggressive, gentle or nervous. Several styles may be perceived in one action. Often there is a gradual change from one style to another, such as from a lazy to an energetic walk. Styles can be characterized by physical adjectives (e.g. fast or slow) and emotional expressions (sadly, aggressively, etc.). In natural language we may describe absolute styles with phrases such as 'slow movement' or 'walking like Mick Jagger', and relative styles by comparative forms such as 'more aggressive'.

Automated identification of styles is possible by associating verbal descriptions with recorded example motions, which in turn are represented by numerical features. An absolute style can be represented as a collection of individual example motions and modeled as a statistical distribution. Analogously, a relative style can be represented as a collection of motion pairs showing differences in that style, and the distribution of differences can be modeled as a vector in feature space [27].

The main contribution of this paper is a new and more accurate method for constructing vector based definitions of relative styles. The basic idea is for each style to find the essential features that in all examples unanimously change when the amount of perceived style changes, and to ignore other features. To accomplish this we need systematical acting of example motions, perceptual annotation of the styles, and individual feature selection for each style.

We also present an implemented system for controlling parametric motion synthesis with the style definitions. The control is indirect as we automatically generate variations of a motion and evaluate which variation shows the desired style best, and then change the synthesis parameters accordingly. Therefore, the style control is independent of the synthesis method. Furthermore, we show with user tests that the produced style vectors accurately predict perceptual evaluations of styles and that the style definitions generalize from one actor to others. Promising results have been achieved with relative styles fast, slow, aggressive, lazy, excited, energetic, calm, limping, healthy, depressed and busy.

We limit our practical experiments to human locomotion, such as walking or running, characterized by physical adjectives and emotional expressions. However, the method for evaluating style is not limited to locomotion and may be extended to non-cyclic motions. We leave out symbolic aspects of conversational gestures that require knowledge of a specific culture to be correctly understood. However, the manner how gestures, such as waving a fist, are performed could still be controlled with our method.

In the following sections, we first review previous work on motion style. Then we present our method, detailed by calculation of low level motion features, creation of the style definitions, and the style-based control of motion synthesis. Finally, a study is described on how well the style definitions and motions produced by style-controlled interpolation synthesis match human perceptions. We conclude with limitations and potential extensions of the method.

2 Related Work

In this section, we review techniques for editing style in captured motion, and studies on the perception and verbal description of styles. Based on these, we discuss how style semantics and low-level motion synthesis methods have been matched.

2.1 Motion Style in Computer Animation

Traditional motion capture does not separate style from action but the motion is replayed as it is. Only space-time constraints necessary for retargeting the motion to a different character are imposed [7]. All stylistic variations are performed by the real actor. If needed for later use, they are stored in a database and then selected by indexing with a style attribute [12].

One way to approach style explicitly is to model it as the difference between a specific and a regular action. As two captured motions seldom are in the exactly same phase, warping in space and time is usually needed to make them comparable. Hsu et al [8] used machine learning to construct a linear time-invariant model with example motion pairs to model the stylistic difference between the motions. The model enables transforming new neutrally acted motions to the learned style in real-time.

When changing motion styles, we do not always need to have a specific motion sample as a target. Instead, we can try out how editing low level motion data affects the perceived styles. Bruderlin and Williams [2] proposed equalization in frequency space as a tool, demonstrating for example calm and nervous movement resulting from low and high pass filtering, respectively. Min and Chai [14] developed a generative graph model for motion synthesis, separating finite structural variations for content actions (such as walking) and continuous style related variations (such as walking speed and step size). In these works expressive style is not modeled explicitly, and thus cannot be controlled directly.

Yet another approach is to model a motion signal as a sum of editable components. This allows both analysis of various important features, and synthesis as recombination of components. Fourier spectrum edited by filtering [2] is one example of this type of modeling. Alternatively, action sequences can be statistically modeled as combinations of base functions produced with Principal Component Analysis (PCA) [20, 23] or Independent

Component Analysis (ICA) [18]. With PCA and additional statistical analysis, emotional and gender-related styles such as nervous, sad, relaxed, male and female have been successfully identified [20, 21].

Our method was inspired by these works. Particularly, we adopted from Bruderlin and Williams [2] the frequency components as motion features.

2.2 Perception and Verbal Description of Style

One of the earliest systems enabling semantic control of motion style identified verbs as distinct actions and adverbs as versions of the actions in different styles [17]. We basically follow this, although the distinction is not strict. Some verbs include a stylistic aspect, against which adverbs tend to be relative modifications. For example, scuffing may imply dampened motion, and slow running may be almost the same as fast walking, and still all these are variations of the same action of locomotion. There are also complex interactions between different styles conceptualized as adverbs, as one tends to imply another. For example, the perceived gender of a moving character can be affected by the perceived amount of anger and sadness [10].

Many methods exist for recognition of actions based on groups of individual examples [16]. As absolute style can be represented by individual examples, the same methods could be applied. However, we concentrate on relative style as that allows precise iterative fine tuning of styles.

In a recent study about natural language in describing human motion, verbal annotations of motion samples were related to their calculated low level features such as distances between body parts, velocities, accelerations and absolute positions [6]. Plotting the results against PCA components of the features (Fig. 1) indicates that verbs tend to be localized in partly overlapping clusters, whereas adverbs are less unanimously annotated (Fig. 2). This coincides with the intuitive understanding that unlike verbs that can be used alone, adverbs are linguistic modifiers that tend to reflect as directions rather than locations in the feature space. Fine control of style with absolute definitions would require dividing all verb clusters to smaller pieces such as slow walking, aggressive walking and sad walking. This would require a lot more samples and annotations than defining only generic actions.

Motion style in dancing can be described with Laban notation, based on expert terms related to effort and shape of motions [3]. The definitions of expert terms need to be learned explicitly, while natural language does not require additional training. Also, the terms

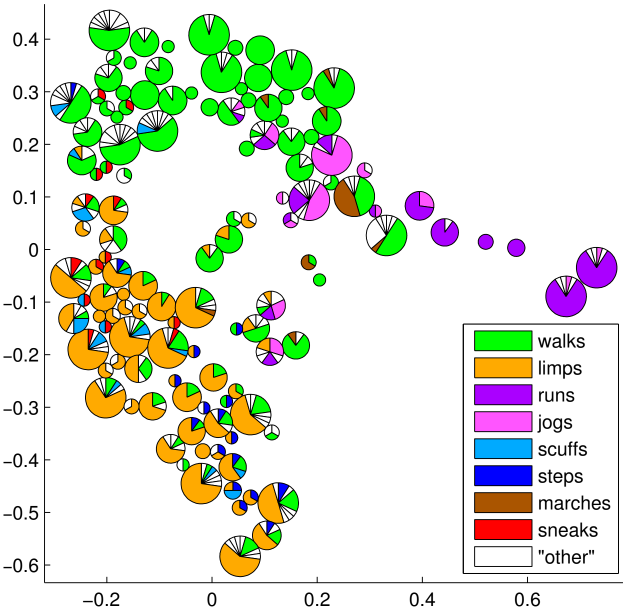


Fig. 1 Motions, annotated with verbs, mapped on the first and second normalized PCA components of numerical motion features [6]. The surface area of the pies is proportional to the number of annotations and the distances between the pies reflect the similarity of the motions

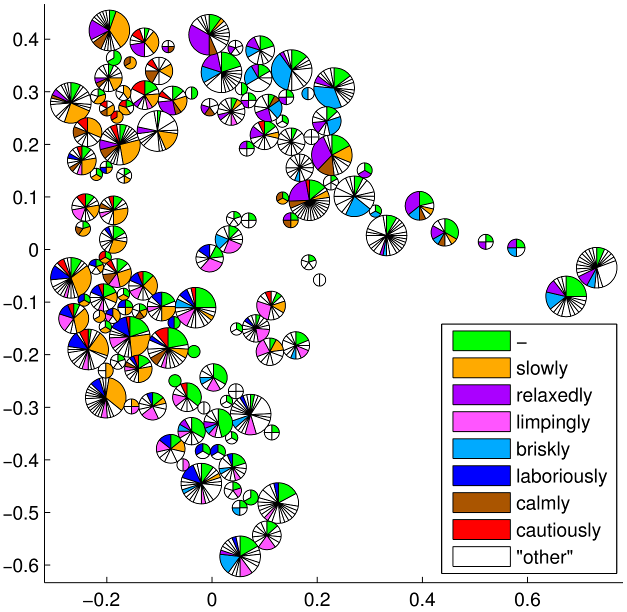


Fig. 2 Same motions as in Fig. 1, annotated with adverbs, mapped on the first and second normalized PCA components of numerical motion features [6]. (”-” colored with green means that no adverb was given.)

that have precise definitions inside dancing may not be sensible in other motion categories. This is why we use laymen terms instead of expert definitions.

Psychological studies on perception and recognition of affects in bodily motion mostly assume discrete non-overlapping classes of emotions or use abstract affective dimensions instead of natural language [11]. For this reason they are not directly applicable to controlling continuous motion in animation. Important considerations for psychologists are whether acted or authentic emotions should be studied and if the ground truth comes from actors or observers [11]. These questions are more straightforward in animation as synthetic virtual characters do not have real intentions or feelings. What counts is observers' perception alone.

For animators, the stimuli used in studies about emotions in human motion may look very simplified, often consisting only of point-lights on a black background [10,21]. One reason for simplified appearance is that giving too many details, such as facial expressions, may divert attention away from the motion or modulate the perception of emotions [1,5]. In our work, we compromise and use a stick figure. It lacks details but helps in perceiving postural differences between motions.

Our general goal is to allow animators to adjust style of synthesized motion by words in natural language. In earlier research, this approach has been taken with action commands (verbs) such as 'walk five steps and pick up the object' [14]. An alternative non-verbal approach has been to sketch key-poses of a motion sequence, allowing more precise positioning of the actions but still lacking control over other style related attributes [26]. In this paper, we focus on refining the actions by relative commands such as 'do the same, but more slowly and sadly'.

2.3 Matching Style Semantics and Synthesis

Given a verbal description, a corresponding motion can be produced in different ways. A rich database of motion samples acted in all possible styles would be easy to use but impractical to generate. More viable is a parametric model, mapping verbal instructions to navigation in the parameter space of a synthesis engine.

Motion interpolation is a parametric method that can produce a continuous range of styles between compatible original samples [17]. However, the results of interpolation cannot be accurately predicted from the parameters and verbal descriptions of the original samples when styles are mixed. For example, interpolation between sad and aggressive motions could end up looking neutral or showing sadness in the pose and aggression in accelerations.

Although motion inaccuracy, such as foot sliding, is a problem in goal-oriented actions, it can be alleviated by sophisticated interpolation methods [15]. The same has not been possible with styles. In lack of automatic evaluation, manual annotation of several interpolated samples is necessary to make reliable predictions, and the number of possible interpolations grows combinatorially with the number of new original motions.

Modeling motion styles with a functional decomposition (PCA or ICA) allows direct synthesis by recombination of the desired components [20,18,23]. These methods offer orthogonal parameters which can be tuned independently to reach a desired style. However, the parameters do not automatically match with natural language descriptions of styles that may be partially synonyms or opposites. Every parameter can affect several perceived styles depending on how the styles were correlated in the original motions used in calculating the components. For example, adjusting emotional styles described with phrases such as sadness or relaxedness can also affect styles related to the body shape of the character [20]. Another problem is that although these methods enable extrapolation of motion from one sample to new situations, such as a different speed, extrapolation carries a risk of producing motions that are not physically realistic if not used carefully. For reasons stated above, we think component based methods are not suitable for describing relative styles with natural language.

Treating motion signals as frequency bands is another candidate for style synthesis [2,22]. For example, Bruderlin and Williams [2] report that amplifying high frequencies can add "a nervous twitch" to a walking motion. However, this may not be the case for all input motions. Assigning meaning to the parameters may be even more difficult than with interpolation or component based methods, as the frequencies may have different effects depending on the input motion.

2.4 Vector Based Style Definitions

Relative differences in style between motions can be modeled in a numerical feature space as style vectors representing the direction of increasing perceived style. Zhuang et al. [27] defined style vectors statistically as differences between means of motion samples performed by an actor repetitively in different styles (Fig. 3a). Styles of new motions can then be compared by calculating their difference in projection onto the style vector (Fig. 3b).

The idea of style vectors is to provide a numerical measure for relative style differences which in turn can

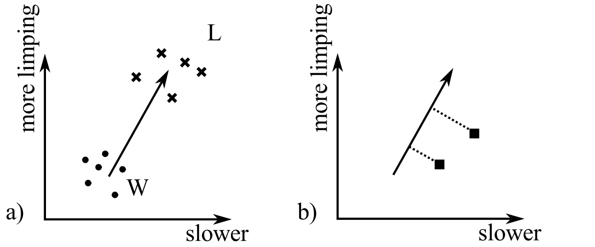


Fig. 3 a) a training set of walking (W) and limping (L) motions, with the style vector that points toward the learned limping style. b) styles of two motions compared by projecting them onto the style vector

be used for iteratively adjusting motion synthesis parameters towards a desired style. Previously a natural language description for the style vectors was more of an afterthought, and the descriptions were not validated in practice [27]. In our work, we follow the same principle, but consider an accurate match between numerical and linguistic descriptions to be vital for a usable style definition. This pushed us to develop the method further.

3 Catching the Essence of a Style

Our aim is to let animators control motion styles by computational features. But how do we know which of them are relevant for a style? As some styles are related to postures and others to limb velocities, the same set of features is not relevant for all.

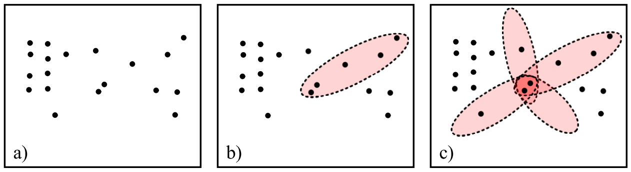


Fig. 4 Features used to model styles. a) the set of all computed features b) the features that correlate in stereotypically acted examples, c) subsets formed by acting the same style in different ways - the essential features are in the intersection

In the set of all potential features (Fig. 4a), we want to identify those relevant for each particular style. To find them, we may ask an actor to perform motions in varying intensities and calculate which features consistently change when the style gets stronger (Fig. 4b).

However, if some features correlate with multiple styles, they cannot make a distinction between those. For example, the style vector in Figure 3 would judge a slower but otherwise normal walk as limping, because limping typically is a slow motion.

We propose a solution where an actor performs variations of one style combined with other simultaneous styles instead of just repetitions of one style (for example, 'sad+fast' and 'sad+aggressive' in addition to plain 'sad'). This way we can identify the essential features common to all cases where a style difference appears (intersection of ellipses in Fig. 4c).

As a lot of irrelevant features may get dropped out with this refinement, our approach requires the number of original features to be high in order to ensure that at least some essential features can be found.

4 Motion Synthesis with Refined Style Vectors

Below we present a method for calculating style vectors that more accurately identify different styles. We first describe the feature set used for evaluating styles in motion, and then give details on how style vectors are constructed from acting a set of sample motions through perceptual annotation to calculation of the vectors (Fig. 5).

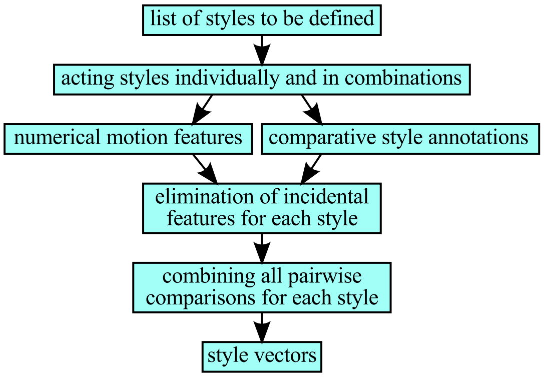

list of styles to be defined
acting styles individually and in combinations
numerical motion features
comparative style annotations
elimination of incidental features for each style
combining all pairwise comparisons for each style
style vectors

Fig. 5 Overall process of creating style vectors

We also present a system for controlling interpolation based motion synthesis by style vectors. Motion control starts from one sample that presents the desired action in any style. An animator can then use natural language based descriptors to adjust the motion towards the desired style while keeping action the same.

Our method does not rely on mapping one style to one synthesis parameter. Instead, we build style control as gradual navigation in the parameter space by solving a parameter combination that best produces a desired change in style. Virtually any synthesis method can be used, as we treat motion synthesis as a black box, containing possible post-processing steps such as inverse kinematics.

4.1 Motion Features

Computational comparison of motions requires numerical motion features. Our aim is to find generic features that can be used for detecting style in the data captured from any type of human movement. Raw motion capture data consists of time varying signals with values for each frame. From that, we calculate a set of features where each value represents a short motion segment (approximately 2-10 seconds long) as styles are partly dynamic properties which cannot be seen in single frames. In order to accurately identify many styles, we need a lot of potential features, out of which a suitable subset is defined for each style.

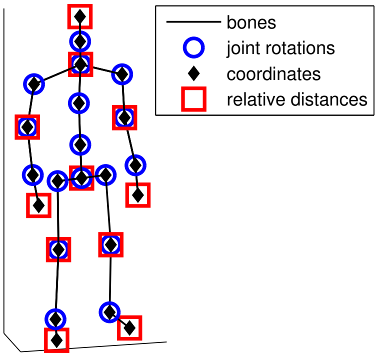


Fig. 6 Skeleton structure of the motion capture data showing bones of constant length, the 18 rotating joints, and the 22 points of body, the coordinates of which are used in calculation of our features, either as such or as 55 relative distances (see Table 1)

We model the human body as a hierarchical skeleton structure (Fig. 6) with constant bone lengths and joint rotations that vary in each frame. The lowest level of per frame data includes coordinates, velocities, accelerations and rotations at joints (expressed as quaternions). From the velocities we take both absolute values and the components along axes of the character's local coordinate system. Also, we include all pairwise distances between pelvis, neck, head, elbows, hands, knees and feet. This set of motion signals has been useful in recognition of action verbs [6].

To expand the set to be more suitable for motion style, we also calculate how the signals vary in frequency domain [21]. Following the filtering method by Bruderlin and Williams [2], we divide the initial signals into seven frequency bands. From the original signal captured with 100 Hz sample rate we extract approximately the ranges 0.1–0.5–1.1–2.2–4.5–9–18–50 Hz. Thus, we have 301 motion signals (Table 1) in eight versions (original and the seven frequency bands) making 2408 signals altogether. To summarize the signals as numbers that describe whole motion segments, we take their means and standard deviations over all frames in the segment. With this the number of dimensions in our feature space becomes 4816.

As similar movements can be performed using the left or the right side of the body, we consider these to be of identical style. To make the 4816 features the same in both cases, we first checked which of them already are mirror invariant. For the rest, taking absolute values makes equal the features directly related to sideways motion. Instead of velocities for the left and the right hand, we sort them pairwise to get velocities of the slower hand and the faster hand. For features related to sideways motion of paired limbs, we multiply the value of one side with -1 and then apply the sorting.

To make our features invariant of body size, we divide all coordinate values by the height of the actor, which also scales velocities and accelerations to comparative ranges. Other normalizations between actors are not applied as they could harm the identification of styles related to bodily structures.

4.2 Creating Vector Based Style Definitions

For defining styles, we asked an actor to perform a regular walk and eight style variations relative to that: fast, slow, relaxed, tense, angry, sad, limping, and excited. We also asked the actor to perform combinations of two styles (all except fast+slow and relaxed+tense as those styles can be considered mutually exclusive), making altogether 35 motions. Our amateur actor was able to perform the style combinations with noticeable variation, although it required him to consciously analyze different aspects of individual styles and devise a way to combine them.

To obtain perceived differences in styles, motions were annotated in pairs using a questionnaire, where each page displayed two video samples for visual comparison. The annotator provided up to three words to describe their differences and quantified them on a scale *'a little/somewhat/a lot more'*. To avoid biased answers, we did not offer any predefined choices for the words.

For avoiding excessive manual work, we limited annotation to the motion pairs that most likely show differences between styles. As depicted in Figure 7, we included those where a double style (i.e. actor instructed to present two styles simultaneously) is compared with a regular motion (26 pairs), and all possible combinations of single styles (56 pairs). The latter comparison was done in both ways (shown separately as swapped pairs) to encourage the annotator to name opposite differences. Altogether this reduced our questionnaire

Table 1 The motion signals from which the features representing motion segments are derived

Signals	Consisting of	Number of dimensions
Positions in local coordinate system of the character	22 body parts with 3 values each (The center of pelvis is the root joint for which only elevation coordinate is taken into account.)	64
Absolute velocities	22 body parts with 1 value each	22
Velocities along local coordinate axes	22 body parts with 3 values each	66
Absolute accelerations	22 body parts with 1 value each	22
Distances between pelvis, neck, head, elbows, hands, knees and feet	pairwise combinations of the 11 body parts with 1 value each	55
Joint rotations as quaternions	18 joints with 4 channels each (The base of neck and shoulders has three overlapping joints. The three bones starting from the central hip represent the pelvis and share the same rotation.)	72
Total		301

from 1190 possible comparisons to only 82. The annotation was done by one of the authors. To ensure that the results are not biased, we later made a validation by crowdsourcing.

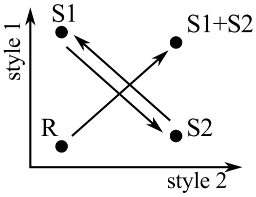


Fig. 7 Motion pairs used in the comparative annotation: double styles (S1+S2) against regular motion (R), and single styles both ways against each other (S1 and S2)

From the annotation data, we selected 13 most common verbal descriptions that appeared in at least five example pairs: fast, slow, aggressive, lazy, excited, energetic, calm, limping, healthy, depressed, busy, relaxed and tense. For these styles we proceeded to calculate style vectors. Eighteen other verbal descriptions appeared in the annotation data less than five times.

For each style we collected the results of annotation in form of Table 2, with one row for each pairwise comparison where the style was seen (N varying from 5 to 25 depending on how many motion pairs got labeled with the style). The vector $\mathbf{c_x}$ consists of the differences of numerical features between the compared motions. The perceived style difference a_x is a value scaled from '*a little/somewhat/a lot more*' to 1, 2 or 3 respectively. In the last column A_x is the sum of all difference values given in the comparison for any styles. In our case, as the motion pairs were shown only once during the questionnaire and the annotator may give at most three styles per motion pair, the maximum value for A was 9.

From this table we identify those features that agree in all comparisons, i.e. we select those y for which $c_{x,y}$ has the same sign in all rows x=1...N. These are the essential features for recognizing the particular style. The other features, which are incidental, we eliminate from the style vector, thus effectively reducing dimensionality of the feature space. However, as the essential features are not the same for all styles, we retain all original features, only weighting them for this style by multipliers defined as:

Table 2 Summary of collected data for one annotated style, with a row for each pairwise comparison in the questionnaire.

Vector of feature differences in the displayed motions	Perceived style difference	Sum of all perceived differences
$\mathbf{c}_1 = \langle c_{1,1}, c_{1,2}, ..., c_{1,4816} \rangle$	a_1	A_1
$\mathbf{c}_2 = \langle c_{2,1}, c_{2,2}, ..., c_{2,4816} \rangle$	a_2	A_2
...	...	...
$\mathbf{c}_N = \langle c_{N,1}, c_{N,2}, ..., c_{N,4816} \rangle$	a_N	A_N

$$\begin{aligned} m_y &= 1 \text{ if } \forall x : c_{x,y} \geq 0 \lor \forall x : c_{x,y} \leq 0 \\ m_y &= 0 \text{ if } \exists x : c_{x,y} > 0 \land \exists x : c_{x,y} < 0 \end{aligned} \tag{1}$$

The multipliers are then used to create eliminated versions $\mathbf{s}_x$ of the difference vectors $\mathbf{c}_x$:

$$\mathbf{s}_x = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{4816} \end{bmatrix} \mathbf{c}_x \tag{2}$$

The final style vector $\mathbf{u}$ is formed as a weighted mean of the reduced difference vectors

$$\mathbf{u} = \frac{1}{N} \cdot \sum_{x=1}^{N} \left(\left(\frac{a_x}{A_x} \right)^2 \cdot a_x \cdot \mathbf{s}_x \right) \tag{3}$$

where we use the style difference a_x as weight, normalized by its proportion of all style differences given in that comparison (A_x). The previous method [27] treated all motion examples equally and used an unweighted mean as a style vector. We explicitly try to utilize the variance in motions and emphasize the features that contribute to a style. Therefore, we give more weight to the comparisons where the amount of the annotated style is large (multiplication with a_x) and makes up a large percentage of all styles (squared term).

4.3 Vector Based Control of Motion Synthesis

Out of many possible synthesis methods, we selected motion interpolation as it is widely used in animation software and games for producing blends between motions. Also, interpolation is less prone to unnatural results than those methods that extrapolate outside the range of recorded examples.

The parameters of interpolation tell how much the end result should resemble each input motion. To avoid extrapolation, the parameters must be non-negative and sum to 100%. As the input motions for the interpolation, we took the same 35 locomotions with varying styles that were used in creating the style vectors. We matched the times when the feet get on and off the ground. After time warping, root positions of the character were interpolated linearly. For joint rotations, normalized linear interpolation (*nlerp*) of quaternions [19] was applied. Acceleration spikes that may appear as side effects of time warping were smoothed in a post-processing step. Note that time warping was needed for the interpolation synthesis only. For evaluation of styles – the essential part of our method – it is sufficient that the motions contain the same actions; even the number of cyclic repetitions could vary.

What is an optimal control method for motion synthesis depends on the predictability and cost of synthesizing individual motions. A brute force approach would be to produce style variations randomly and pick one closest to the desired style. Instead, we evaluate the effect of offsetting each synthesis parameter individually and then solve the best combination of changes to the parameters, effectively performing a gradient search.

Motion interpolation is a locally stable synthesis method, meaning that adding a small offset to a parameter has a small predictable effect on the produced motion. Then we can model the effects of parametric changes linearly with a Jacobian matrix:

$$\mathbf{J}\mathbf{x} = \mathbf{u} \tag{4}$$

where each column of $\mathbf{J}_k$ is the vector of partial changes in feature values $\mathbf{u}$ caused by changing the corresponding parameter x_k alone.

Looking for a desired style change $\mathbf{u}$, the required parameter change can be found by solving this equation for $\mathbf{x}$. An exact solution is unlikely as the number of parameters is much lower than the number of motion features. Pseudoinverse is a suggested solution in inverse kinematics, but we used an off-the-shelf least squares solver (*lsqnonneg* in Matlab) as it easily finds an approximate solution with minimal error while keeping the synthesis parameters inside the interpolation range. The steps for finding new synthesis parameters with a desired style are listed in Algorithm 1.

Algorithm 1 Finding new synthesis parameters

```
Start with arbitrary parameters ($param$) that produce the desired action
while user not satisfied do
    User selects a desired style change (style vector $u$)
    $J = \textsc{ConstructJacobianAt}(param)$
    $x = \textsc{SolveLinearSystem}(J, u)$
    $x = \textsc{ScaleToLimitMaximumParameterChange}(x)$
    $param = param + x$
    $param = \textsc{ScaleTo100Percent}(param)$
    $\textsc{SynthesizeAnimationWith}(param)$
end while
```

We built a user interface (shown in Online Resource 1) for trying out the style control in practice. It shows an animation of the current motion and allows relative adjustment towards a desired style. The user may control either the desired change in each style and let the algorithm tune the parameters, or adjust the 35 synthesis parameters directly. In our experience, the latter was more tedious especially when trying to simultaneously get more than one style visible in the motion.

Examples of motions produced with our system are shown in Figure 8 and as animations in Online Resource 1. They show how aspects such as step size, velocities, posture, and limb trajectories behave when the style is changed. The trajectories show that excess feet sliding did not appear even though inverse kinematics was not used.

The visualization method in Figure 8 appears to be a novel technique for presenting motion style in still images. In our opinion it shows the dynamics of motion better than a series of stick figures.

Our implementation in a multicore computer is fast enough for interactive applications. Main part of the computation is spent on synthesizing motion trials for calculating the Jacobian. As the motions are synthesized independently, the task can be distributed to the available cores in parallel.

5 Experimental Validation

In this section, we test how well the style vectors work in practice. In our first experiment the styles seen by human observers in a set of walking motions are compared with automatic evaluations done with style vec-

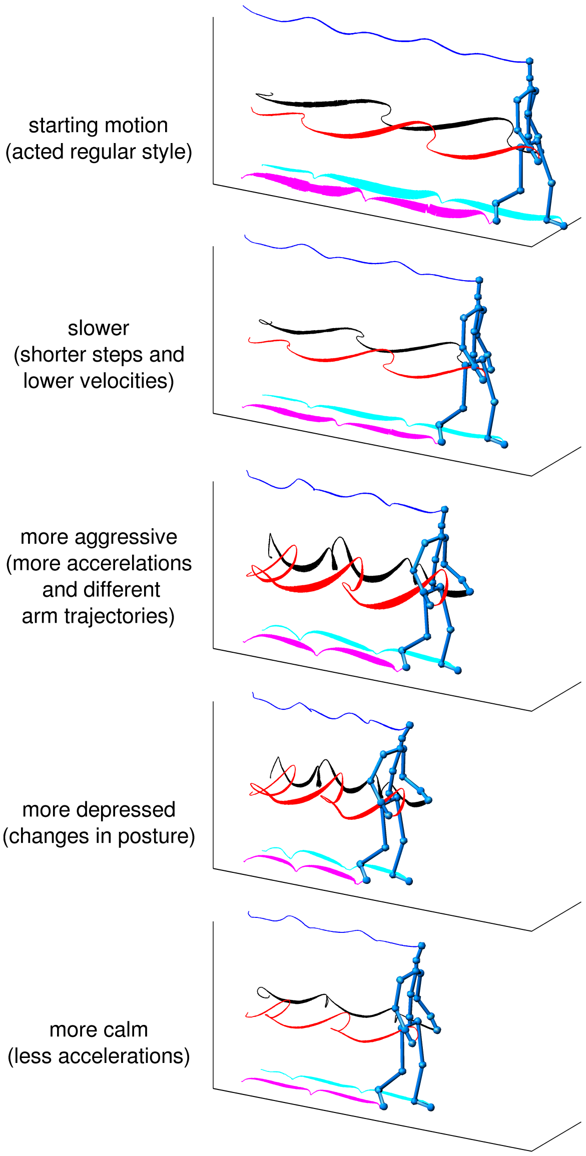


Fig. 8 Control of walking style by relative style commands. Starting from an acted motion on top, each picture shows incremental changes towards the bottom. Trajectories leading to the final pose are shown for head, hands and feet. Line thickness indicates velocity

tors. Next, we assess the impact of our feature elimination process to the quality of style vectors. In a second experiment, we test if human observers recognize style adjustments produced by motion synthesis with our method.

5.1 Validation of Style Definitions

Accurate style vectors should enable automatic evaluation of styles acted by new actors and the result should be agreed on by new human observers. To test this aspect, we produced a new set of locomotions, annotated the perceived style differences with a crowdsourced questionnaire, and compared the annotations to the style evaluations produced with style vectors.

The new set of locomotions was performed by four actors and acted in similar styles as before when creating the style vectors. As some actors were not able to perform all style combinations properly, we enriched the motion set by also creating 50/50% interpolations between the actor's motions, disregarding those where interpolation caused visible artifacts. From this set of 168 unique motions we produced 347 pairs that in our opinion differed in at least one of the annotated styles.

The numbers of motion pairs and unique motions for each style are shown in Table 3. Some pairs were used as examples of several style differences. Also, one motion sample may appear in several pairs.

The motion pairs were shown as stick figure animations on a web page. The observer was given one of the 13 style descriptions and asked to evaluate on a five-point scale (*much more / slightly more / equal amount / slightly less / much less*) how much one sample shows the given style compared with the other. The presentation order was balanced so that each pair of videos was shown twice, with the order of comparison reversed. We ran the questionnaire using the web-based crowdsourcing platform CrowdFlower and received a total of 10569 ratings randomly distributed among 456 participants. For quality control, we included a test in the start that required separating pairs of 100% identical videos from pairs that showed extremes of opposite styles. This way we could be sure that all the participants were at least able to view the videos.

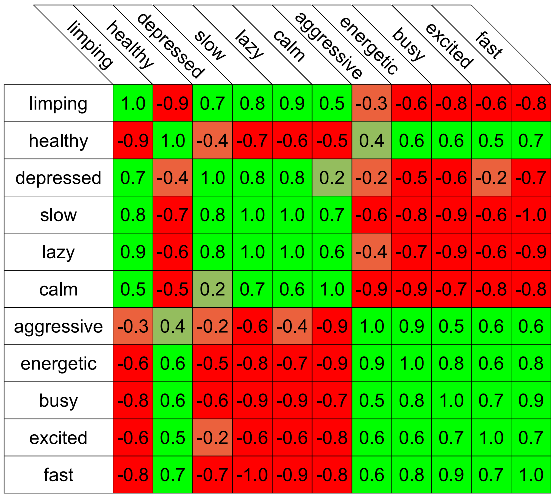
The crowdsourcing service provided us with six or seven ratings for every combination of a style word and a video pair. At this point, we pruned the data by only keeping those combinations in which majority (at least four) of the participants had agreed on which of the videos had more of the mentioned style. This reduced the number of style word and video pair combinations from the original 1041 to 952 which we took as ground truth for our tests.

To measure the accuracy of style definitions we tested how well automatic evaluation of style differences agrees with the ratings of the majority of the questionnaire participants. Style difference of a motion pair was automatically evaluated by calculating the dot product between the style vector and the vector of feature differences between the two motions. The sign of the dot product was taken as indication of which motion shows more style. In this setup, the chance level for accuracy is 50%.

The results, shown in Table 3, tell that most of the style definitions reached at least 90% accuracy and sev-

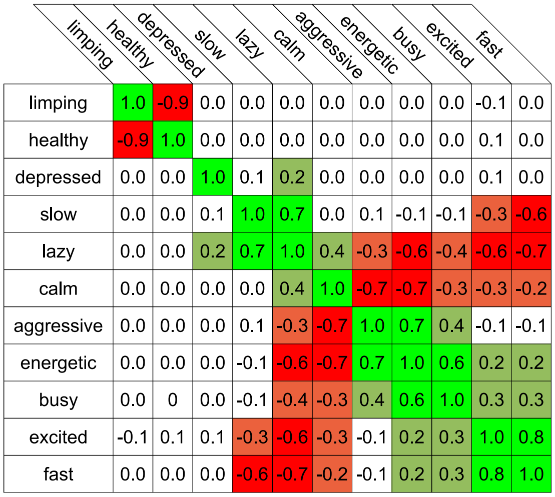
Table 3 Accuracies of automatic style evaluation

Style word	Accuracy %	Number of motion pairs	Number of unique motions
fast	100	170	93
slow	100	166	89
aggressive	100	70	57
lazy	100	62	46
excited	100	28	30
energetic	98.8	80	49
calm	98.5	65	48
limping	97.1	68	57
healthy	96.8	63	47
depressed	92.0	88	46
busy	90.0	20	33
relaxed	77.1	35	39
tense	59.5	37	43



	limping	healthy	depressed	slow	lazy	calm	aggressive	energetic	busy	excited	fast
limping	1.0	-0.9	0.7	0.8	0.9	0.5	-0.3	-0.6	-0.8	-0.6	-0.8
healthy	-0.9	1.0	-0.4	-0.7	-0.6	-0.5	0.4	0.6	0.6	0.5	0.7
depressed	0.7	-0.4	1.0	0.8	0.8	0.2	-0.2	-0.5	-0.6	-0.2	-0.7
slow	0.8	-0.7	0.8	1.0	1.0	0.7	-0.6	-0.8	-0.9	-0.6	-1.0
lazy	0.9	-0.6	0.8	1.0	1.0	0.6	-0.4	-0.7	-0.9	-0.6	-0.9
calm	0.5	-0.5	0.2	0.7	0.6	1.0	-0.9	-0.9	-0.7	-0.8	-0.8
aggressive	-0.3	0.4	-0.2	-0.6	-0.4	-0.9	1.0	0.9	0.5	0.6	0.6
energetic	-0.6	0.6	-0.5	-0.8	-0.7	-0.9	0.9	1.0	0.8	0.6	0.8
busy	-0.8	0.6	-0.6	-0.9	-0.9	-0.7	0.5	0.8	1.0	0.7	0.9
excited	-0.6	0.5	-0.2	-0.6	-0.6	-0.8	0.6	0.6	0.7	1.0	0.7
fast	-0.8	0.7	-0.7	-1.0	-0.9	-0.8	0.6	0.8	0.9	0.7	1.0

Fig. 9 Correlations between style vectors without elimination of incidental features with values greater than 0.15 and less than -0.15 in green and red backgrounds respectively



	limping	healthy	depressed	slow	lazy	calm	aggressive	energetic	busy	excited	fast
limping	1.0	-0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0
healthy	-0.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
depressed	0.0	0.0	1.0	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0
slow	0.0	0.0	0.1	1.0	0.7	0.0	0.1	-0.1	-0.1	-0.3	-0.6
lazy	0.0	0.0	0.2	0.7	1.0	0.4	-0.3	-0.6	-0.4	-0.6	-0.7
calm	0.0	0.0	0.0	0.0	0.4	1.0	-0.7	-0.7	-0.3	-0.3	-0.2
aggressive	0.0	0.0	0.0	0.1	-0.3	-0.7	1.0	0.7	0.4	-0.1	-0.1
energetic	0.0	0.0	0.0	-0.1	-0.6	-0.7	0.7	1.0	0.6	0.2	0.2
busy	0.0	0	0.0	-0.1	-0.4	-0.3	0.4	0.6	1.0	0.3	0.3
excited	-0.1	0.1	0.1	-0.3	-0.6	-0.3	-0.1	0.2	0.3	1.0	0.8
fast	0.0	0.0	0.0	-0.6	-0.7	-0.2	-0.1	0.2	0.3	0.8	1.0

Fig. 10 Correlations between style vectors after the elimination of incidental features

eral even got 100% of the test pairs correct. This is a good result as the style vectors were produced from motions of one actor and annotations of one person, while the evaluation set had four actors and hundreds of observers.

We also observe that the styles relaxed and tense were less accurately defined than the other styles. The use of style vectors for controlling motion synthesis sets an acceptability level for the accuracy. For example, if an animator asks for a more relaxed motion, with 77% accuracy the system would give a more relaxed motion only three times out of four. Therefore we dropped the relaxed and tense style definitions from the rest of the experiments.

5.2 Assessment of the Impact of Incidental Features

The main difference between our method and the previously published one [27] is the elimination of incidental features. If the elimination step works, it should remove false correlations between styles and preserve correlations only when the styles defined are semantically overlapping. In order to evaluate the impact of feature elimination, we calculated pairwise correlations between style vectors produced without elimination (Fig. 9) and compared them with those produced by our elimination process (Fig. 10).

The correlations in Figures 9 and 10 reveal that the elimination step does make the style vectors more independent from each other. For example, before elimination the styles limping and slow have a correlation of 0.8. This means that increasing the amount of visible limping would also increase the amount of slowness (cf. Fig. 3). However, after the elimination step, the styles limping and slow have a correlation that rounds to 0.0 meaning that with these style vectors, adjusting the level of limping can be done without affecting the perceived slowness. The correlations that remain non-zero after the elimination are reasonable as those style pairs can be semantically considered close to synonyms or opposites (such as slow and lazy, or calm vs. energetic).

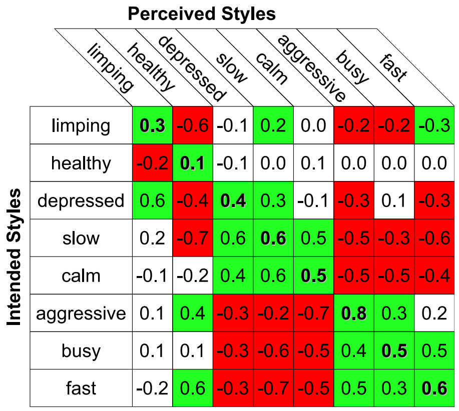
5.3 Validation of Synthesized Styles

The first validation experiment indicated that the style vectors correspond well to human perception when testing with acted motions. This implies that the vectors should allow accurate control of motion synthesis. To directly test it, we ran the first validation experiment again with motions produced by interpolation synthesis.

As starting motions for the test, we took equally spaced samples from the parameter space of the interpolation. Each sample was produced with 50% of one parameter and the remaining 50% equally divided for all others. This gave us 35 initial motions. From every initial motion we created eight modifications, each adjusted to display a fixed amount more of a particular style. This was done by offsetting the parameters by a vector calculated from Eq. 4. To reduce the human evaluation task, we used only the eight style vectors which did not show large positive correlations in Figure 10, *i.e.* limping, healthy, depressed, slow, calm, aggressive, busy and fast. Thus, we ended up with 280 pairs showing an initial motion and its adjusted version.

Perceptual evaluation of the 280 motion pairs was done with a similar questionnaire as in section 5.1. We received a total of 4485 individual ratings randomly distributed among 314 participants.

Answers of the questionnaire were scaled so that ±2 means *much more/less style*, ±1 *slightly more/less style*, and 0 stands for *no change in style*. From this data, the mean scores for every combination of intended and perceived styles were calculated, and statistically significant differences from zero with p-value 0.05 were identified. The means are based on 70 or 71 evaluations. Figure 11 shows a confusion matrix of the results.



Intended Styles	limping	healthy	depressed	slow	calm	aggressive	busy	fast
limping	**0.3**	-0.6	-0.1	0.2	0.0	-0.2	-0.2	-0.3
healthy	-0.2	**0.1**	-0.1	0.0	0.1	0.0	0.0	0.0
depressed	0.6	-0.4	**0.4**	0.3	-0.1	-0.3	0.1	-0.3
slow	0.2	-0.7	0.6	**0.6**	0.5	-0.5	-0.3	-0.6
calm	-0.1	-0.2	0.4	0.6	**0.5**	-0.5	-0.5	-0.4
aggressive	0.1	0.4	-0.3	-0.2	-0.7	**0.8**	0.3	0.2
busy	0.1	0.1	-0.3	-0.6	-0.5	0.4	**0.5**	0.5
fast	-0.2	0.6	-0.3	-0.7	-0.5	0.5	0.3	**0.6**

Perceived Styles
Intended Styles

Fig. 11 Mean scores from evaluation of style adjustments. Scores on white do not statistically differ from zero (p=0.05), significant positive differences are green and significant negative differences red

If controlling the synthesis is successful, the intended style should get a significant increase due to the adjustment and even larger change than any other style. The diagonal of Figure 11 shows that all intended styles were actually perceived to increase. However, the change was not always the largest. This is understandable in cases where the initial motion already has plenty of the intended style visible; then the increase cannot be very large as discussed in the next section.

6 Discussion

Our method for creating style vectors does require some talent and concentration from actors, people instructing the actors and the person annotating the motions. Therefore, the method does not replace the work of animation professionals. However, since the style vectors can be used for controlling styles of new motion sets, the fruits of the labor can be enjoyed by people who are not experts in motion capture techniques.

Our animated demo (Online Resource 1) and the related experiments show that style vectors enable control of several styles simultaneously. How intense the styles eventually get, is up to the acted motions and the synthesis method used. Interpolation limits expressivity to that of the input motions while extrapolation may produce more intense but sometimes unnatural style.

We model a relative style with one style vector, but acknowledge that a global vector is not sufficient in all cases. For example 'natural' is a property that has a maximum from which there are many, even opposite ways to get away, and its negation (unnatural) is ambiguous. Local style vectors that always point to the maximum (or away from it for respective negations) could work better than a global vector. We were able to define 'healthy' with a global style vector as our set of examples had limping as the only unhealthy movement. However, asking an actor to perform in an unhealthy way could provoke a demand for more specific instructions. This may be the reason why the style vector for healthy did not score so well in our experiment (Fig. 11). As most starting motions of the test already looked quite healthy, it could not be improved much.

We acknowledge that low correlations between style vectors (Fig. 10) create expectations of better separation between styles than the results of the crowdsourced experiment imply (Fig. 11). Varying proficiency of the English language among the globally distributed participants may explain part of the overlapping use of style words. To our knowledge, all previous publications presenting style oriented motion synthesis have completely omitted a similar validation. Therefore, our work can be considered state-of-the-art in this respect.

A risk in our method is that a style vector can degenerate to zero if no essential features are left after elimination. This can happen if the style is ill-defined, poorly acted, or annotated inconsistently due to human errors. We do not consider the last reason to be a serious one as style definitions can be created by relatively low amount of annotations by just one person. Therefore, correcting annotation mistakes does not mean much work.

The style vectors could be produced by different means than our process. We considered using Support Vector Machine (SVM) to find a hyperplane separating two style classes and applying its normal as the style vector. SVMs work well in classification of absolute concepts represented with individual examples such as verbs [4,25]. For relative concepts a better option is Ranking SVM [9], but we did not adopt that either as the method by Zhuang *et al.* [27] or our refinement of it are simpler to implement and computationally less intensive.

Our method could be developed further by experimenting with new actions, styles, actors, low-level features and synthesis methods. Preliminary experiments on reusing style definitions with other actions have been promising. For example, definitions for styles slow and aggressive based on locomotion seemed to apply to hand waving or turning. However, trying to make a hand wave more limping created random looking results.

A practical use case for our method is communication with virtual characters. Bodily motions could drastically improve expressivity compared to facial expressions or symbolic messages alone. Another use case is browsing in a motion library. Starting from one motion with the desired action, its variations in style could be found with relative steps instead of having to watch all possible alternatives.

7 Conclusions

In this paper, the semantic meaning of verbally described styles has been grounded in numerical motion data more precisely than before. Our main contribution is the method producing more accurate style vectors by eliminating other features than those essential for recognizing a style.

We have presented a method for indirectly controlling motion synthesis by style words. We let an arbitrary synthesizer generate candidate motions, evaluate them with style vectors, and select the best. For a stable synthesis method, such as interpolation, the desired changes in style can be mapped to offsets in synthesis parameters. Controlling a large number of parameters this way is more user friendly than adjusting them directly.

Our evaluation of the method shows that style definitions created from motions of one actor and annotated by one observer, accurately predict styles observed by other people in motions performed by other actors.

In a practical application, a virtual actor could be first commanded to perform an action such as 'walk' or 'run' and then the performance could be fine-tuned by relative commands such as 'more limping' or 'more aggressively'.

Preliminary results suggest that the method generalizes many styles over motion categories, such as from locomotion to turning in place, but further research is needed to find the precise requirements for successful transfer of style.

Acknowledgements This work has been supported by the HeCSE graduate school and the project Multimodally grounded language technology (254104) funded by the Academy of Finland. The Mocap toolbox by Neil Lawrence [13] was used in this research.

References

1. Aviezer, H., Hassin, R.R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., Bentin, S.: Angry, disgusted, or afraid?: Studies on the malleability of emotion perception. Psychological Science **19**(7), 724–732 (2008)
2. Bruderlin, A., Williams, L.: Motion signal processing. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95, pp. 97–104. ACM, New York, NY, USA (1995)
3. Chi, D., Costa, M., Zhao, L., Badler, N.: The emote model for effort and shape. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00, pp. 173–182. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (2000)
4. Cho, K., Chen, X.: Classifying and visualizing motion capture sequences using deep neural networks. In: Proceedings of the 9th International Conference on Computer Vision Theory and Applications, VISAPP2014 (2014)
5. Clavel, C., Plessier, J., Martin, J.C., Ach, L., Morel, B.: Combining facial and postural expressions of emotions in a virtual character. In: Z. Ruttkay, M. Kipp, A. Nijholt, H. Vilhjálmsson (eds.) Intelligent Virtual Agents, *Lecture Notes in Computer Science*, vol. 5773, pp. 287–300. Springer Berlin Heidelberg (2009)
6. Förger, K., Honkela, T., Takala, T.: Impact of varying vocabularies on controlling motion of a virtual actor. In: R. Aylett, B. Krenn, C. Pelachaud, H. Shimodaira (eds.) Intelligent Virtual Agents, *Lecture Notes in Computer Science*, vol. 8108, pp. 239–248. Springer Berlin Heidelberg (2013)
7. Gleicher, M.: Retargetting motion to new characters. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '98, pp. 33–42. ACM, New York, NY, USA (1998)
8. Hsu, E., Pulli, K., Popović, J.: Style translation for human motion. ACM Trans. Graph. **24**(3), 1082–1089 (2005)
9. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02, pp. 133–142. ACM, New York, NY, USA (2002)
10. Johnson, K.L., McKay, L.S., Pollick, F.E.: He throws like a girl (but only when hes sad): Emotion affects sex-

decoding of biological motion displays. Cognition **119**(2), 265–280 (2011)

11. Kleinsmith, A., Bianchi-Berthouze, N.: Affective body expression perception and recognition: A survey. Affective Computing, IEEE Transactions on **4**(1), 15–33 (2013)
12. Kovar, L., Gleicher, M., Pighin, F.: Motion graphs. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02, pp. 473–482. ACM, New York, NY, USA (2002)
13. Lawrence, N.: Mocap toolbox for matlab. Available on-line at http://staffwww.dcs.shef.ac.uk/people/N.Lawrence/mocap/ (2011)
14. Min, J., Chai, J.: Motion graphs++: A compact generative model for semantic motion analysis and synthesis. ACM Trans. Graph. **31**(6), 153:1–153:12 (2012)
15. Mukai, T., Kuriyama, S.: Geostatistical motion interpolation. In: ACM SIGGRAPH 2005 Papers, SIGGRAPH '05, pp. 1062–1070. ACM, New York, NY, USA (2005)
16. Poppe, R.: A survey on vision-based human action recognition. Image and Vision Computing **28**(6), 976–990 (2010)
17. Rose, C., Cohen, M., Bodenheimer, B.: Verbs and adverbs: Multidimensional motion interpolation. Computer Graphics and Applications, IEEE **18**(5), 32–40 (1998)
18. Shapiro, A., Cao, Y., Faloutsos, P.: Style components. In: Proceedings of Graphics Interface 2006, pp. 33–39. Canadian Information Processing Society (2006)
19. Shoemake, K.: Animating rotation with quaternion curves. SIGGRAPH Comput. Graph. **19**(3), 245–254 (1985)
20. Troje, N.F.: Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of Vision **2**(5), 371–387 (2002)
21. Troje, N.F.: Retrieving information from human movement patterns. Understanding events: How humans see, represent, and act on events pp. 308–334 (2008)
22. Unuma, M., Anjyo, K., Takeuchi, R.: Fourier principles for emotion-based human figure animation. In: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95, pp. 91–96. ACM, New York, NY, USA (1995)
23. Urtasun, R., Glardon, P., Boulic, R., Thalmann, D., Fua, P.: Style-based motion synthesis. Computer Graphics Forum **23**(4), 799–812 (2004)
24. Wang, X., Jia, J., Cai, L.: Affective image adjustment with a single word. The Visual Computer **29**(11), 1121–1133 (2013)
25. Wu, J., Hu, D., Chen, F.: Action recognition by hidden temporal models. The Visual Computer **30**(12), 1395–1404 (2014)
26. Yoo, I., Vanek, J., Nizovtseva, M., Adamo-Villani, N., Benes, B.: Sketching human character animations by composing sequences from large motion database. The Visual Computer **30**(2), 213–227 (2014)
27. Zhuang, Y., Pan, Y., Xiao, J.: A Modern Approach to Intelligent Animation: Theory and Practice, chap. Automatic Synthesis and Editing of Motion Styles, pp. 255–265. Springer (2008)