Emotion from Motion

Kenji Amaya
Dept. of Mechanical and Environmental Informatics
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152, Japan
email:kamaya@mes.titech.ac.jp

Armin Bruderlin
ATR Media Integration & Communication
Seika-cho Soraku-gun
Kyoto 619-02, Japan
email:armin@mic.atr.co.jp

Tom Calvert
Simon Fraser University
Burnaby, B.C., V5A 1S6, Canada
email:tom@cs.sfu.ca
phone: (604) 291-4588 fax: (604) 291-4424

Abstract

This paper introduces a model to generate "emotional" animation from "neutral" human motion. Using techniques from signal processing, our method calculates certain *emotional transforms* which are then applied to existing motions of articulated figures in order to produce the same motions, but with an emotional quality such as angry or sad. These transforms capture the difference between a neutral and emotional movement with respect to two components: *speed* (timing), and *spatial amplitude* (range) of a movement.

Since the transforms are applied as global operations, they provide a convenient and efficient way to adapt motion-captured, simulated or keyframed animation of articulated figures to different situations and characters.

Keywords: human figure animation, motion capture, motion control, digital signal processing.

Introduction

In recent years, human animation has played an increasing role in such areas as advertising, entertainment, education, scientific visualization and simulation. However, while many motion generation methods have been published, human animation is still in its infancy especially in the representation of expression, personality and emotion when compared to real human movement. Much of the difficulty in animating human motion can be attributed to the many degrees of freedom that must be controlled even for simplified models. Another challenge in animating human movement is the fact that humans are very sensitive observers of each others' motion, in the sense that we can easily detect erroneous movement ("it simply doesn't look right"), although we often find it much more difficult to isolate the factor which causes the movement to look incorrect.

Motion capture techniques have come to the rescue since they preserve the distinctive "signature" of the real movement. However, motion capture has the disadvantage that special equipment is required and current systems allow for only limited editing capabilities to adapt a movement once it is captured; this requires the whole data capture process to be repeated if a motion sequence slightly different from an already captured one is desired.

In this paper, a method to produce emotional animation from neutral, expressionless motion is proposed. This method can be divided into two parts: identification of certain *emotional transforms* by signal processing techniques and application to a "neutral" human motion to generate the same movement, but with an emotional trait. For example, if we apply the "angry" transform T_A to an animated sequence of a person drinking from a cup, the result will be an angry person drinking.

Our approach modifies existing animation data of articulated figures and therefore makes the use of motion-capture, procedural, physically-based and keyframe techniques more meaningful and useful. By applying elementary techniques from signal processing, a high-level interface to producing emotional animation is achieved. This approach is related to several other research efforts: Unuma et al. [1, 2] apply Fourier transformations to data on human walking for animation purposes. Through Fourier expansions of the joint angles, a basic 'walking' factor and a 'qualitative' factor like "brisk" or "fast" are extracted. These factors are then used to generate new movements by interpolation and extrapolation in the frequency domain, such that now a walk can be changed continuously from normal to brisk walking, or a walk can be changed smoothly into a run. Litwinowicz uses recursive filters to produce "lag, drag, and wiggle" effects to keyframed two-dimensional animated motion in a system called Inkwell [4]. Bruderlin and Williams [5] introduce a number of signal processing techniques for human figure animation which support various animation effects such as applying multiresolution filtering to exaggerate a movement, or automatically aligning two movements in time via non-linear timewarping. They also apply displacement mapping to conveniently edit motion-captured data. A similar technique called motion warping was proposed by Witkin and Popovic [6].

For the purpose of animating "emotions"[1], an emotion is considered to be a kind of *secondary* movement which piggybacks on top of a *primary* movement (see [10] for a discussion on primary/secondary movement, and [11] for a categorization for gestures). In this paper, we are mainly concerned with body movement, neglecting facial expression and speech which are also important factors in convincingly animating emotions in human characters. However, as explained in the next section, our technique is general in the sense that no matter what emotional motion is provided in calculating the emotional transform, the "difference" between the emotional and neutral motion is convincingly applied to a new, neutral movement.

In Section , our approach to generate emotional animation is described in more detail. Section discusses how the emotional transforms are derived and applied to new neutral movement. Results and examples are given in Section , and conclusions and future work are addressed in Section .

Basic Approach

Our technique to animate human emotions involves the following steps:

1. capture the motion of human subjects performed with different emotions, such as angry, sad, neutral;
2. for each emotion, calculate an *emotional transform* which is the "difference" between the neutral and emotional movement;
3. apply this emotional transform to a new, neutral movement.

We used an optical motion capture system (OPTOTRAK [12]) to record the movements as shown in Figure 1. In an experimental setup, subjects were asked to perform two motions with different emotions as well as in a neutral manner: "pick up the glass of water, drink from it, and put it back onto the table", and "knock at the door three times". A script was presented to each subject to provide situational context to each emotion[2]. Each emotion was recorded three times. Six infrared emitting diodes (IRED's) were attached to the subjects' head, shoulder, elbow, wrist and hand. The system tracked the locations of each IRED[3] by three calibrated cameras while calculating the absolute positions in space over time. From the six positions of the IRED's, nine rotational degrees of freedom (joint angles) were calculated in order to animate a human figure: two for sternum, three for the shoulder, one for the elbow and three for the wrist.

Steps two and three above address the main focus of this paper: how to abstract and represent the difference between a neutral and an emotional captured motion, so that it can be applied to a new movement to make it emotional. After careful analysis of the motion-captured data, we identified two components which vary noticeably over the various emotions: *speed* (timing) and *spatial amplitude* (range) of the motion. Figure 2 illustrates these variations in

[1] Whereas the study of human emotions has lead to some consensus over a definition about what constitutes an emotion — environmental and psychological events influence brain processes that actively modulate clearly observable behaviors [7] — several models exist on how to classify emotions into angriness, sadness, happyness, fear, etc. [7, 8]. Research also indicates that there are significant cultural differences in how emotions are expressed and perceived [9].

[2] A total of ten emotions or moods were captured — neutral, angry, sad, happy, fearful, tired, strong, weak, excited and relaxed — although for the analysis here we concentrate on just neutral, angry and sad.

[3] A sampling rate of 120 Hz was chosen; accuracy was within 0.5mm.

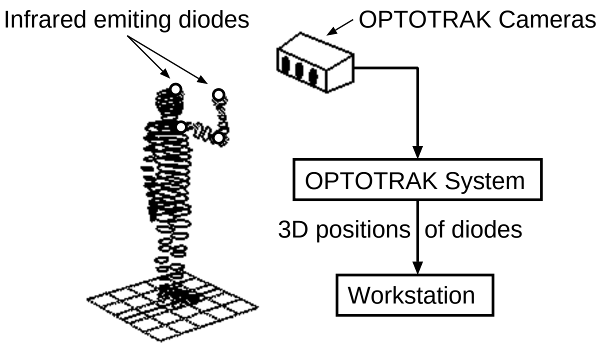


Figure 1: Motion capture system.

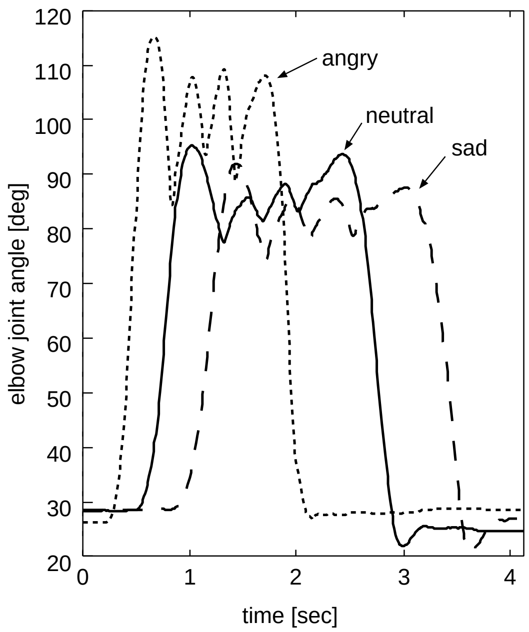


Figure 2: Knocking with different emotions.

the motion-captured data of the knocking movement; the joint angles of the elbow show significant differences in time and amplitude for the neutral, angry and sad motion, respectively.

As we explain in the next section, the *speed* transform has been implemented as a non-linear time-warping technique, and the *spatial amplitude* transform is based on signal amplifying methods. Applying these transforms produces an emotional human movement from a new, neutral movement. To verify our method, we applied the following procedure, examples of which are presented in section :

1. calculate the angry and sad transforms from captured drinking data.

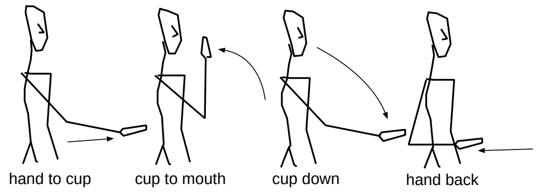


Figure 3: Basic periods for drinking motion.

2. apply the angry and sad transforms to neutral, captured drinking data and compare with captured angry and sad data.

3. apply the angry and sad transforms (calculated from drinking data) to neutral knocking data and compare with captured angry and sad knocking data.

Algorithm

This section focuses on how the transforms for speed (section) and spatial amplitude () are derived from existing motion-captured data, and then applied to a new neutral movement. In order to provide a general technique which works with different motion data, we first subdivide both the neutral and emotional captured data into units of motion, called "basic periods". These periods are bounded by the time when the velocity of the wrist change its direction, or they lie between the extrema which separate extension and flexion in a joint [4]. For example, for the drinking motion we used to derive the emotional transforms the basic periods are "hand to a cup, cup to mouth, cup down, hand back" as shown in Figure 3.

Transform of Speed

The first step to obtain the *speed* transform is to calculate the absolute speed of the end effector which is the wrist point in our case for the drinking motion (where speed is defined along the trajectory of the wrist) for both neutral and emotional motion data (see Figure 4).

After determining the basic periods as defined above, one of them is selected and integrated along the trajectory. For the drinking motion the period with the longest duration was selected ("cup down"). The calculated data can now be represented as:

$$s = f_N(t) = \int_0^t |\mathbf{v}_N(\tau)| d\tau \; ; \qquad (1)$$

[4] The basic periods can also be specified directly by the user based on the joint angle or velocity trajectories.

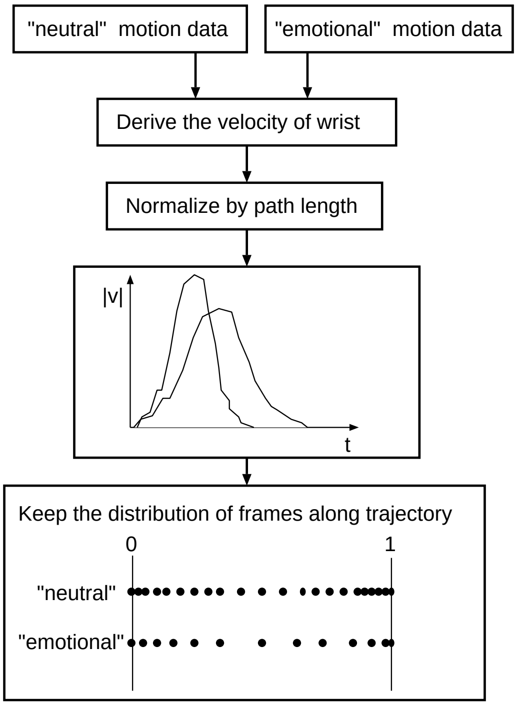


Figure 4: Algorithm to obtain speed transform.

$$s = f_E(t) = \int_0^t |\mathbf{v}_E(\tau)| d\tau \; ; \qquad (2)$$

where t is the time, s the position along the trajectory, $\mathbf{v}(\mathbf{t})$ the 3D velocity vector of the wrist and the subscript N and E denote "neutral" and "emotional ", respectively. These data are normalized along the trajectory as follows, where t_{end} is the duration of the basic period:

$$\hat{s} = \hat{f}_{N_{or}E}(t) = f_{N_{or}E}(t) / \int_0^{t_{end}} |\mathbf{v}_{N_{or}E}(\tau)| d\tau \; . \qquad (3)$$

The distribution of frames $\rho(\hat{s})$ is calculated by

$$\rho_{N_{or}E}(\hat{s}) = n\frac{dt}{d\hat{s}} = \frac{n}{\hat{f}'_{N_{or}E}(\hat{f}^{-1}(\hat{s}))} \; ; \qquad (4)$$

where n is the number of frames per second; $\rho_N(\hat{s})$ and $\rho_E(\hat{s})$ are used as templates when applying the transform.

The first step in applying the speed transform to a new, neutral movement is the calculation of the distribution of its frames, followed by a division into "basic periods" as illustrated in Figure 6.

For each basic period, the following calculation steps are then performed as shown in Figure 5: normalize (scale) period in length; substitute the "emotional" distribution of the frames for the neutral one;

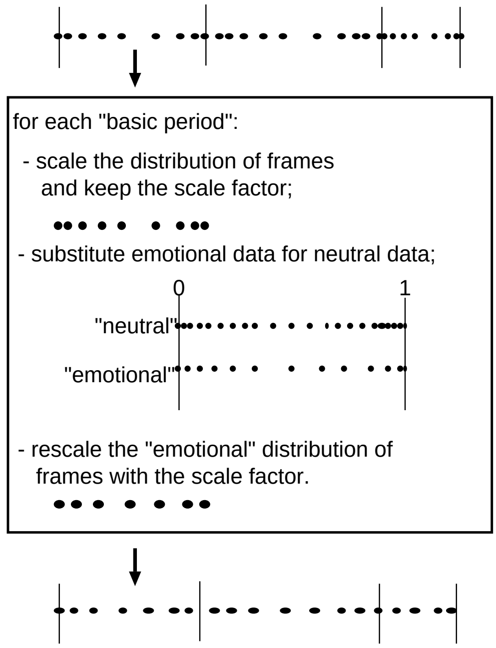


Figure 5: Application of speed transform.

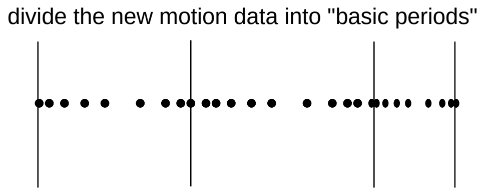


Figure 6: Keyframe distribution of new motion data.

re-scale basic period (inverse of initial normalization).

This warped frame distribution is used as a correspondence table in the "emotional" joint angle calculations. Figure 7 gives an example: the emotional joint angle of the 5th frame is obtained by interpolating between the 6th and 7th frame in the original data, because the emotional 5th frame corresponds to the 6.8th frame in the original data.

Transform of Spatial Amplitude

The transform of spatial amplitude is obtained by applying the algorithm described below:

for both neutral and angry motion, divide the joints into four categories corresponding to the levels of hierarchy of the articulated figure. This division is necessary because the range of motion for the joints are substantially different in each category. Figure 8 shows which joint belongs to which category. For

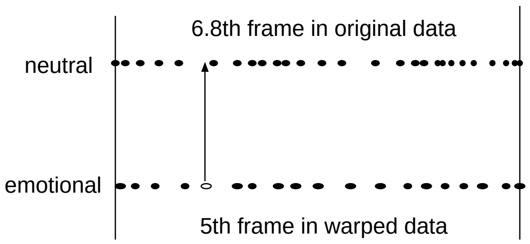


Figure 7: Generation of joint angle data.

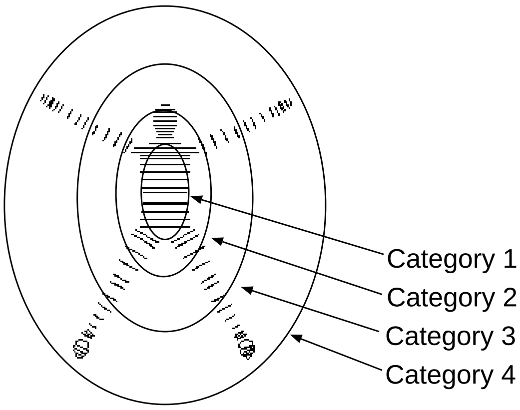


Figure 8: Joint categories.

example, the sternum belongs to the category one, and a knee belongs to category three[5].

Each category represents a multi-dimensional space which is defined by its joint angles and time t. Let us define these multi-dimensional spaces by $\boldsymbol{\theta}^i_{N_{or}E}(t)$, where i $(1 \leq i \leq 4)$ denotes the corresponding category. For example, the multi-dimensional space for category one has 35 degrees of freedom (34 joint angles plus time).

In order to extract the intensity of the spatial amplitude from both "neutral" and "emotional" motion data, the factor $d^i_{N_{or}E}$ is defined as follows for each basic period in turn:

$$d^i_{N_{or}E} = max(|\boldsymbol{\theta}^i_{N_{or}E}(t) - \{\boldsymbol{\theta}^i_{N_{or}E}(t_{init})(1-t) + \boldsymbol{\theta}^i_{N_{or}E}(t_{end})t\}|) \; ; \qquad (5)$$

where $|\cdots|$ is the Euclidean norm operation. This calculation can be represented conceptually as shown in Figure 9. A straight line is drawn from the initial point to the end point of the current basic period for each category i for both "neutral" and "emotional" motion data. The maximum distances $d^i_{N_{or}E}$ between this straight line and the trajectory of motion are calculated in this space, and can be considered as intensities of spatial amplitude for each neutral and emotional motion.

[5] Category one involves 34 joints; category two involves 26 joints; category three involves four joints; category four involves 12 joints in our model of the human figure.

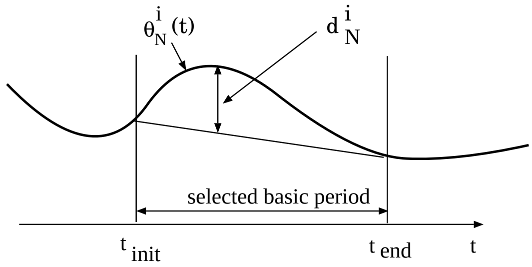


Figure 9: Intensity factor of spatial amplitude d^i.

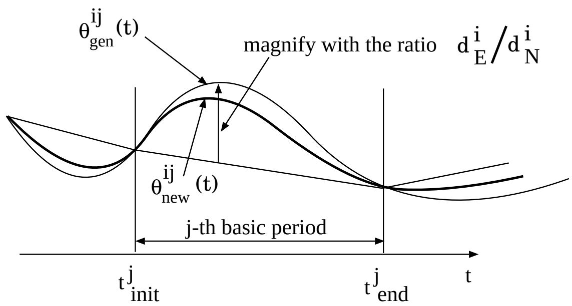


Figure 10: Transformation of spatial amplitude.

The spatial amplitude transform is now applied to a new, neutral motion, $\boldsymbol{\theta}^{ij}_{org}(t)$, where j is the corresponding basic period. This results in a new, emotional motion, $\boldsymbol{\theta}^{ij}_{gen}(t)$, defined by

$$\boldsymbol{\theta}^{ij}_{gen}(t) = \frac{d^i_E}{d^i_N}\boldsymbol{\theta}^{ij}_{org}(t) + \frac{d^i_N - d^i_E}{d^i_N}\{\boldsymbol{\theta}^{ij}_{N_{or}E}(t^j_{init})(1-t) + \boldsymbol{\theta}^{ij}_{N_{or}E}(t^j_{end})t\} \; . (6)$$

The basic idea of this amplitude transform is nonlinear-magnification (see Figure 10); draw a straight line from the initial point to the end point of each basic period in the time-joint angles space of each category. Magnification is performed on the distance between this straight line and the trajectory of motion by the ratio $\frac{d^i_E}{d^i_N}$ for each category i and each basic period j.

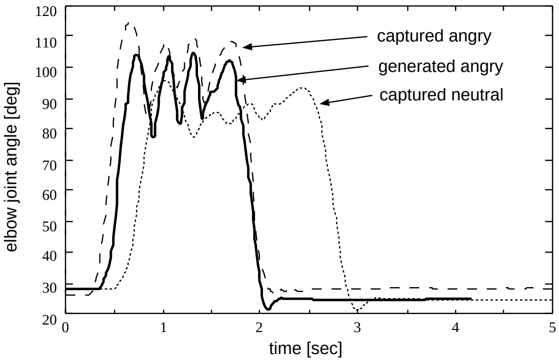


Figure 11: Comparison between generated and captured elbow joint-angles.

Results

Examples

After applying the angry and sad transforms derived from the motion captured drinking motion to the neutral, captured drinking data, we obtained a close match with the "real" (motion-captured) angry and sad drinking data. The same emotional transforms were then applied to the motion of knocking at a door. The resulting motions were compared to the motion-captured angry and sad knocking movements to verify the model. Figure 11 shows the elbow joint angle during a knocking movement; it is seen that the generated angry movement data is a good fit to the real captured data.

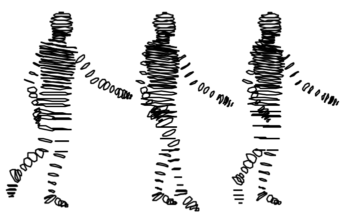
The same emotional transforms have also been applied to a plain, keyframed kicking motion which produced believable emotional variations of the motion. Figure 12 shows snapshots of this kicking animation.

High Frequencies

Besides the speed and amplitude transforms, we searched for other components in the joint angle signals which could produce a significant effect between neutral and emotional signals. Initially, the frequency content was also considered as an important aspect of movement. However, results of frequency analysis revealed that high frequency components in the data are neglectable in both the "emotional" and "neutral" signals. Figure 13 shows the power spectrum of the wrist position; we can see that the signals do not have major components above 10 Hz.

Thus, high frequencies in human motion can be ignored for the purposes of animation where, in any event, a sampling rate of 30 frames/sec or less does

animated "neutral" kicking motion;



generated "sad" kicking motion;

generated "angry" kicking motion;

Figure 12: Keyframes for animated kicking motion.

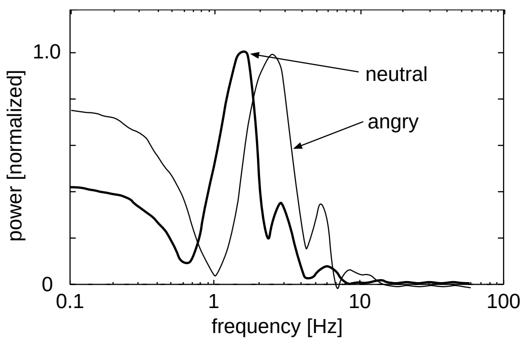


Figure 13: Power spectrum of wrist position data.

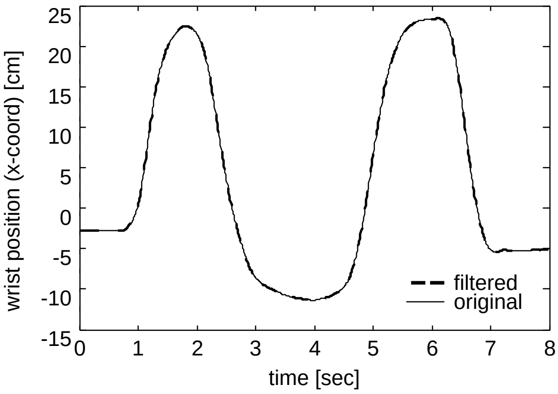


Figure 14: Comparison between low-pass filtered data and original data.

not allow for accurate reproduction of higher frequency components. Furthermore, what we can not *see* in animation is not important: Figure 14 gives a comparison between low-pass filtered data (applying a Gaussian filter kernel of width five) and original data plotted against time; there is no noticeable difference between the two signals.

Phase Shift

Rather than high frequencies which are often regarded as providing the realistic "signature" in human animation, we believe that movement *phase shift* is a distinctive feature of real human motion. Phase shift is the amount with which the movements of the individual joints overlap[6]. Figures 15 and 16 show the joint angle velocity over time for the neutral and angry drinking data, where each signal is normalized by its maximum value. The maximum velocity timings show this phase shift clearly. Moreover, we think that these phase shifts are different between "neutral" and "angry" motion. More research will be necessary to determine how such a phase shift transform could be incorporated into our algorithm.

Conclusions and Future work

A model has been developed to produce "emotional" animations from "neutral" human motion. The method is based on signal processing techniques which analyze experimental data of emotional human motion and extract the difference between emotional and neutral movement. Two components are isolated to define this difference, a *speed* and a *spatial amplitude* transform. These two components are

[6] This has long been regarded as an important principle in articulated figure animation [3].

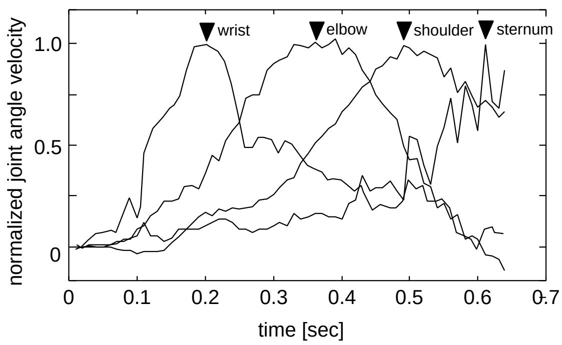


Figure 15: Neutral joint angles.

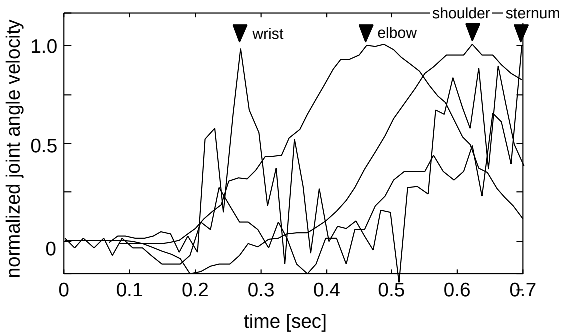


Figure 16: Angry joint angles.

then applied to new, neutral movements to generate "emotional" movements. In order to verify our the approach, angry and sad transforms from a motion-captured drinking sequence have been applied to a neutral knocking motion. The calculated angry and sad knocking motions came very close to the real, captured angry and sad knocking motions. By establishing various categories for degrees of freedom of the human figure, this method is general with respect to employing an emotional transform derived, say, for the arm to a motion of another body part. Emotional kicking motions were generated by applying the drinking transform to the lower body.

By automating the generation of emotional animations our technique facilitates the reuse and adaptation of existing motions of articulated figures and makes the use of motion libraries of keyframed, motion-captured, simulated or procedurally generated movements more meaningful. Also, since the computations for deriving the emotional transforms are done off-line, this approach could well be useful in changing the emotions of virtual actors on the fly in real-time environments.

We are currently investigating how this approach can be extended in various ways. Besides emotions, it is desirable to generate animations with different personalities, cultures and genders. Also, we hope to capture the difference in motion between an old man and a young boy in a similar way. Furthermore, using the same emotional transforms derived from human drinking data to produce an angry walking sequence of a dog, for instance, from the neutral motion would be a practical generalization of our method.

References

[1] Munetoshi Unuma and Ryozo Takeuchi, "Generation of human motion with emotion", in *Computer Animation '93, Proceedings*, 1993, pp. 77–88.

[2] Munetoshi Unuma, Ken Anjou, and Ryuzo Takeuchi, "Fourier principles for emotion-based human figure animation", in *Computer Graphics (SIGGRAPH '95 Proceedings, Los Angels, CA, August 6-11)*, 1995, pp. 91–96.

[3] John Lasseter, "Principles of traditional animation applied to 3–D computer animation", in *Computer Graphics (SIGGRAPH '87 Proceedings)*, 1987, vol. 21, pp. 35–44.

[4] Peter Litwinowicz, "Inkwell: A 2 1/2–D animation system", in *Computer Graphics (SIGGRAPH '91 Proceedings)*, 1991, vol. 25, pp. 113–122.

[5] Armin Bruderlin and Lance Williams, "Motion signal processing", in *Computer Graphics (SIGGRAPH '95 Proceedings, Los Angels, CA, August 6-11)*, 1995, pp. 97–104.

[6] Andrew Witkin and Zoran Popovic, "Motion warping", in *Computer Graphics (SIGGRAPH '95 Proceedings, Los Angels, CA, August 6-11)*, 1995, pp. 105–108.

[7] Robert Plutcheck and Henry Kellersman, Eds., *Emotion: Theory, Research and Experience, Volume 1-3*, Academic Press, 1986.

[8] Carroll E. Izard, Ed., *The Psychology of Emotions*, Plenum Press, New York, 1991.

[9] Shunya Sogon and Makoto Masutani, "Identification of emotion from body movements: A cross-cultural study of Americans and Japanese", *Psychological Reports*, , no. 65, pp. 35–46, 1989.

[10] Claudia Morawetz, "A high-level approach to the animation of human secondary movement", Master's thesis, Simon Fraser University, School of Computing Science, April 1989.

[11] Ray L. Birdwhistell, *Kinesics and Context: Essays on Body Movement Communication*, University of Pennsylvania Press, Philadelphia, 1970.

[12] Northern Digital Inc., *Northern Digital Optotrak Manual and Product Information*, J.R. Krist, 403 Albert Street Waterloo, Ontario N2L 3V2 Canada, October 1991.