May 2000

Interpreting Movement Manner

Liwei Zhao
University of Pennsylvania

Monica Costa
University of Pennsylvania

Norman I. Badler
University of Pennsylvania, badler@seas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/hms

Recommended Citation

Publisher URL: http://dx.doi.org/10.1109/CA.2000.889052

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/hms/47
For more information, please contact repository@pobox.upenn.edu.
Interpreting Movement Manner

Abstract
We describe a new paradigm in which a user can produce a wide range of expressive, natural-looking movements of animated characters by specifying their manners and attitudes with natural language verbs and adverbs. A natural language interpreter, a Parameterized Action Representation (PAR), and an expressive motion engine (EMOTE) are designed to bridge the gap between natural language instructions issued by the user and expressive movements carried out by the animated characters. By allowing users to customize basic movements with natural language terms to support individualized expressions, our approach may eventually lead to the automatic generation of expressive movements from speech text, a storyboard script, or a behavioral simulation.

Keywords
laban movement analysis, gesture, interactive computer animation, natural language control

Comments

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.
Interpreting Movement Manner

Liwei Zhao†, Monica Costa† ‡, and Norm I. Badler†

† Center for Human Modeling and Simulation
University of Pennsylvania, Philadelphia, PA 19104-6389 USA
‡ Computer Science Department
Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

We describe a new paradigm in which a user can produce a wide range of expressive, natural-looking movements of animated characters by specifying their manners and attitudes with natural language verbs and adverbs. A natural language interpreter, a Parameterized Action Representation (PAR), and an expressive motion engine (EMOTE) are designed to bridge the gap between natural language instructions issued by the user and expressive movements carried out by the animated characters. By allowing users to customize basic movements with natural language terms to support individualized expressions, our approach may eventually lead to the automatic generation of expressive movements from speech text, a storyboard script, or a behavioral simulation.

Keywords: Laban Movement Analysis, Gesture, Interactive Computer Animation, Natural Language Control

1 Introduction

In recent years increasing attention has been paid to natural language control of interactive computer animation and natural movement generation of animated characters. Researchers try to build an animation system in which a user can talk to characters using natural language instructions to direct or modify their actions. The animated characters basically understand a small set of motion concepts and stage directions that we use in conversing with other people. Such a system offers a more abstract and higher level of control in which animators manipulate their characters by issuing navigation commands [12], or describing rules of behavior, emotional state, or reactions to situations [5]. However, these systems frequently suffer from a common problem: the low-level movements of the animated characters lack variations to create different expressions and manners. Indeed, the various characters often move in a fairly mechanical manner and share very similar motions.

On the other hand, research in natural movement generation has been continuously making progress in modifying existing motion capture or animated sequences to produce different expressions [19, 22, 1, 11, 25, 21]. They seem to have captured a wide variation in low-level movements. However, the process of leveraging existing motions is often lengthy, requiring users to either generate some function models and characteristic functions, or manually adjust some low-level, non-intuitive parameter settings.

We introduce a new paradigm that combines the advantages of natural language control and natural movement generation — the ability to produce a wide range of natural-looking movements by specifying manners and attitudes with high-level natural language instructions. We take a deep look into movement science with the hope to have a better understanding of the qualitative
aspects of movements and how they vary across individuals. Our EMOTE approach [14, 7] to bridge the gap between manually and procedurally animated characters establishes a new layer of control in which expressiveness is represented by a small number of qualitative parameters. This layer of control gives rise to another layer where characters controlled by natural language instructions show different performances according to adverbs that convey manners and attitudes.

Our approach involves modifying the Effort factors (Space, Weight, Time, and Flow) and the Shape dimensions (Horizontal, Vertical, Sagittal, and Shape Flow) of actions (verbs) using adverbs as modifiers. All the verbs available to the user have a corresponding motion template which is stored in a persistent motion database and defined as key time and pose information as well as some particular Effort and Shape settings. The adverbs in the textual input given by the user are first parsed out by a natural language parser and then translated by a natural language translator into Effort and Shape quality modifiers, which are further applied to the motion templates to vary their performances so that different manners, inner states, personalities, and emotions can be depicted by animated characters.

In the next section, we present some related work. In the following sections, we briefly describe the EMOTE module and the PAR module, which is used to make the connection between natural language instructions and virtual agents. We also describe the whole system architecture where all the modules are integrated together. Then, three series of animation examples are presented to show how to use adverbs to modify the Effort and Shape settings of actions as to exhibit different nuances in the expressions of the underlying movements. Finally, we conclude with the main contributions of our current work.

2 Related Work

Several research projects have been dedicated to the use of natural language based interfaces to interact with animated characters. The AnimNL project [2, 23] is one of the most sophisticated works so far on the use of natural language processing (NLP) in the context of 3D animation. In this project natural language commands convey desired motion and spatial aspects of behavior of the animated characters. In the Improv project, Perlin and Goldberg [20] describe an approach in which actors follow “English-style” scripts. Scripts are sets of author-defined rules governing behavior of actors, which are used to determine the appropriate animated actions to be performed at any given time. Cavazza and Palmer [12] implement an integrated parser based on a linguistic formalism tailored to the processing of the specific natural language instructions required to control a game player character. The parser outputs structured messages to the animation layer which further interprets these messages to generate appropriate behaviors.

The idea of altering existing animation to produce different characteristics is not new. Several researchers have specifically addressed the issue of generating movements with expressiveness. Perlin uses stochastic noise functions to give the “visual impression of personality” to animated puppets [19]. The user can vary the expression depicted by characters by adding a random component to their joints, modifying the bias on their joints, and varying the transition times for different actions. These methods give characters a dynamic presence and a random sense of attractiveness, but do not necessarily present a natural, human-like demeanor. Unuma, Anjyo, and Takeuchi use a continuous rescaled Fourier technique to allow smooth transitions between two captured motions using interpolation, as well as to generate exaggerated motions using extrapolation [22]. They showed examples of walking and running with the following emotions: hollow, vacant, graceful, cold, brisk, happy, vivid and hot.

Amaya, Brudelirin, and Calvert present a more general method for adding emotion to motions. They derived emotional transforms from motion capture data by quantifying the differences be-
between neutral and emotion-driven actions using the speed of the end-effector and the spatial amplitude of joint angle signals [1]. They then use the emotional transforms to add emotions to neutral actions. Further, they are able to apply emotional transform derived from one part of body to another part of body. For example, an emotional transform derived from angry and sad arm movements can be applied to legs to generate angry and sad kicking motions. The authors use the technique to capture ten emotions (or moods): neutral, sad, happy, fearful, tired, strong, weak, excited, and relaxed. Our approach distinguishes from these last two techniques on at least two aspects: first, we do not apply characteristics of one motion to another, but instead assume that each motion has default Effort and Shape settings and use adverbs as modifiers to amplify or diminish different aspects of these settings; secondly, our method is not based on the frequency domain and therefore can deal with noncyclic motions that these two methods failed to capture.

By treating motion parameters as sampled signals, Bruderlin and Williams apply techniques from signal processing to modify motions [11]. Witkin and Popović have a similar system for editing captured or keyframed animation by warping motion parameters curves [25]. Wiley and Hahn [24] produce new motions using linear interpolation on a set of example motions, but their method requires computation and storage exponential in the number of parameters.

Rose, Cohen, and Bodenheimer present a method to leverage existing motion capture or animated sequences that allow real-time interpolation based on the settings of “adverbs” [21]. Using an off-line authoring system, they parameterize the motion “verbs” with “adverbs” and create a verb graph to specify transitions between verbs. At runtime, these structures allow the user to modify animations in real-time by changing adverb settings. Their work is closely related to our own, but we take a very different approach. The “adverbs” they use are not actually real adverbs in the sense of natural language but instead some low-level interpolation parameters. As we mentioned before, our work is based on our EMOTE model, which allows the animation of characters with natural-looking movements through the use of a small number of high-level parameters that represent qualitative aspects of movements. We add one more layer on top of EMOTE to parse out adverbs and translate them into modifiers which in turn modify the high-level Effort and Shape parameters. As a result, users do not have to worry about the low-level interpolation parameters, and characters controlled by natural language commands show different performances according to the user-input adverbs that convey manners and attitudes.

3 EMOTE

Our approach in generating natural movements involves building computational models of a particularly important system called Laban Movement Analysis (LMA). LMA has four major components — Body, Space, Shape, and Effort. The components of LMA that we cover are Effort and Shape [9, 10, 17]. In her PhD dissertation, Chi created and implemented a kinematic analog to the Effort part of this movement notational system [14, 6]. In recent work, we have extended Chi’s implementation of arm Effort movements to include the Shape qualities, plus the torso and legs [13]. We call this system EMOTE (Expressive MOTion Engine).

Effort comprises four motion factors: Space, Weight, Time, and Flow. Each motion factor is a continuum between two extremes: (1) indulging in the quality and (2) fighting against the quality. These extreme Effort Elements are seen as basic, “irreducible” qualities, meaning they are the smallest units of change in an observed movement. Table 1 shows the LMA Effort elements — the extremes for each motion factor.

The Shape dimensions in LMA are Horizontal, Vertical, Sagittal, and Shape Flow. The terms

\[ \text{Shape Flow is primarily concerned with changes of relationship with the body parts and these changes can be} \]
<table>
<thead>
<tr>
<th>Effort</th>
<th>Indulging</th>
<th>Fighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space</td>
<td>Indirect</td>
<td>Direct</td>
</tr>
<tr>
<td>Weight</td>
<td>Light</td>
<td>Strong</td>
</tr>
<tr>
<td>Time</td>
<td>Sustained</td>
<td>Sudden</td>
</tr>
<tr>
<td>Flow</td>
<td>Free</td>
<td>Bound</td>
</tr>
</tbody>
</table>

Table 1: Effort Elements

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Shape</th>
<th>Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical</td>
<td>Rising</td>
<td>Weight-Light</td>
</tr>
<tr>
<td></td>
<td>Sinking</td>
<td>Weight-Strong</td>
</tr>
<tr>
<td>Horizontal</td>
<td>Enclosing</td>
<td>Space-Direct</td>
</tr>
<tr>
<td></td>
<td>Spreading</td>
<td>Space-Indirect</td>
</tr>
<tr>
<td>Sagittal</td>
<td>Retreating</td>
<td>Time-Sudden</td>
</tr>
<tr>
<td></td>
<td>Advancing</td>
<td>Time-Sustained</td>
</tr>
</tbody>
</table>

Table 2: Effort and Shape Affinities

used to describe the extreme attitudes towards these dimensions are Spreading and Enclosing, Rising and Sinking, Advancing and Retreating, Opening and Closing, respectively. In general, Shape changes occur in affinities with corresponding Efforts (Table 2 [10]). Although EMOTE allows independent control of Effort and Shape, the affinities should normally be respected.

Effort and Shape qualities are expressed using numeric parameters that can vary along distinct scales. Each dimension of Effort and Shape is associated with a scale ranging from −1 to +1. The extreme values in these scales correspond to extreme attitudes towards their respective dimensions. For example, a +1 value in Effort’s Weight factor corresponds to a Strong movement; a −1 value in Shape’s Horizontal dimension corresponds to an Enclosing movement. Effort parameters are translated into low-level movement parameters, while Shape parameters are used to modify key pose information.

Our EMOTE system can produce a wide range of expressive, natural-looking movements by utilizing the qualitative aspects provided by Effort and Shape and procedurally converting the qualitative into the non-intuitive quantitative parameters. Our Effort and Shape model is computationally efficient, performing in real-time, and allowing for interactive motion generation and editing. Due to space limitation we refer you to [13] for more technical details.

4 PAR

The Parameterized Action Representation (PAR) [4, 5, 8] is designed to bridge the gap between natural language instructions and the virtual agents who are to carry them out. PAR is therefore constructed based jointly on implemented motion capabilities of virtual agents and linguistic requirements for instruction interpretation. PAR includes slots for many types of information that can sometimes occur linguistically as adjuncts to the main verb phrase. These slots include spatio-temporal information, as well as applicability, preparatory, and terminating conditions. Manner information is encoded which is frequently expressed as adverbs (quickly, gently, carefully, sadly, tiredly, haphazardly). This rich representational structure allows PAR to capture the physical or performance attributes of movements and actions. In our current system, we focus on the manner manipulations. On the NLP level the natural language interpreter, consisting of a parser and a translator, maps the user-input adverbs to some manner modifiers in term of Effort and Shape settings. The modifiers are then combined with the default manner defined in the motion template to supply a new manner in an instantiated PAR (iPAR). On the PAR level, the iPAR with the modified manner is put on to the multi-layered iPAR queue in the corresponding agent process (see section 5). When the iPAR is actually executed after various necessary condition checkings, the manner is passed to the EMOTE level for generating the low-level motion parameters, which essentially govern the movement performance.
5 Natural Language Interpreter

The natural language interpreter consists of a parser and a translator. The parser takes a natural language instruction and outputs a tree identifying its different components such as noun, verb, adverb, preposition, etc. We are using as the basis of our parser the XTAG Synchronous Tree Adjoining Grammar system in C++ [18].

The translator uses the output of the parser as indices to look up a translation table, which is built manually to convert adverbs to appropriate Effort and Shape settings. In the following we show a sample translation table.

**TOP-LEVEL**

```
play / PLAY / MORE_TIME_MODIFIER SPACE WEIGHT FLOW
play / PLAY / TIME_MODIFIER SPACE WEIGHT FLOW
play / PLAY / SPACE WEIGHT TIME FLOW
```

**SPACE**

```
[directly] / (0.2, _, _, _)
[indirectly] / (-0.2, _, _, _)
- / (0.0, _, _, _)
```

**WEIGHT**

```
[strongly] / (_, 0.2, _, _)
[lightly] / (_, -0.2, _, _)
- / (_, 0.0, _, _)
```

**TIME**

```
[slowly] / (_, _, -0.2, _)
[quickly] / (_, _, 0.2, _)
- / (_, _, 0.0, _)
```

**FLOW**

```
[freely] / (_, _, 0.0, _)
[boundly] / (_, _, -0.2, _)
- / (_, _, -0.0, _)
```

**TIME_MODIFIER**

```
[more [slowly]] / (_, _, 0.3, _)
[more [quickly]] / (_, _, 0.3, _)
```

**MORE_TIME_MODIFIER**

```
[more [even]] [slowly] / (_, _, 0.4, _)
[more [even]] [quickly] / (_, _, 0.4, _)
```

According to the translation table, instructions “play slowly,” “play more slowly,” “play even more slowly” are translated into PLAY(0.0, -0.2, 0.0, 0.0), PLAY(0.0, -0.3, 0.0, 0.0), and PLAY(0.0, -0.4, 0.0, 0.0), respectively. (Note: the parameter values are set arbitrarily.) Some adverbs may translate into parameters associated with more than one dimension. For example, “carefully” might translate into Light and slightly Sustained (a Weight value of -0.5 and a Time value of -0.2); “haphazardly” might translate into Indirect and somewhat Free (a Space value of -0.4 and an Effort Flow value of -0.5). This can be similarly done by constructing a more sophisticated translation table. It is important to note that the output of the translator is not the final Effort and Shape settings being applied to the movements, but instead manner modifiers that need to be combined with the default manner associated with a specific verb. As we mentioned previously, each verb is associated with default manners that are defined in terms of Effort and Shape. Some verbs that represent actions which carry no particular manners to be performed may have all of their associated Effort and Shape parameters set to zero. An example is the verb “get”. Other verbs such as punch, press, slash, and dab, on the other hand, imply some particular manners and therefore Effort and Shape settings. In such cases, the final Effort and Shape settings are computed as follows:

\[P_i(verb+adverb) = P_i(verb) + P_i(adverb)\]

where \(P_i\) represents the parameter setting in each Effort and Shape component, and \(|P_i(verb)|\) is the absolute value of the verb setting.

6 System Architecture

Figure 1 shows the architecture of the whole integrated system. We briefly describe the relevant modules for the context of this work.

- **Database:** The motion templates specified through key time and pose information, as well as the default manners defined as Effort and Shape parameter settings, are stored in a persistent database. The keyframes could be defined by the user, or could be formed by an external process, for example, a procedurally
generated motion, or motion captured from live performance.

- **Agent Process:** Each instance of an agent is controlled by a separate agent process, which maintains a queue of all iPARs (instantiated PAR) to be executed by the agent. When the applicability and preparatory conditions are satisfied and the terminating condition is false, the agent process is ready to execute the action according to the execution steps specified in the iPAR [5].

- **GestureNet:** It is responsible for the synchronization of ArmTorsoNet and HandNet which respectively control the arm and torso and hands movements. GestureNet, ArmTorsoNet and HandNet are implemented using PaT-Nets (Parallel Transition Networks) [3]. A PaT-Net is a simultaneously executing finite state automata in which the nodes are associated with actions and connections between the nodes are associated with transition conditions. PaT-Nets make it very easy to wait on completion of actions before moving onto the next action, to execute actions in parallel or in sequence, and to dynamically extend the action structure by invoking other PaT-Nets from nodes of the current one [4, 26].

- **Jack Toolkit:** We use the EAI/Transom Jack toolkit [16] to maintain and control the actual geometry, scene graphs, and human behaviors and constraints.

7 **Animation Examples**

To demonstrate the idea of using natural language instructions with verbs and adverbs to leverage natural movements, we have created a series of animations shown on the accompanying video. All the examples were generated in real-time. The first series of animations focus on using adverbs to gradually modify the Effort Time factor, which in turn affects the dynamics of movements. For example, instructions "sign quickly", "sign more quickly", and "sign even more quickly" are used to incrementally speed up the performance of the American Sign Language (ASL) sentence "I
Figure 2: Sign with Neutral Effort Time (first two rows) and Sign very Quickly (last row)

Figure 3: Play Pianissimo (the first row) and Play Forte (the second row) on the Piano
was sick, but I am well now” (see Figure 2). Also we can make the signs be performed slower and slower using instructions "sign slowly", "sign more slowly", and "sign even more slowly." To show why EMOTE is more versatile than methods that are only working on stretching or compressing a timing curve, in the second series of animations we concentrate on using adverbs to modify the Effort Weight factor. For example, the instruction "play forte" on the piano produces a Strong manner toward the impact of body weight as opposed to a Light manner produced by the instruction "play pianissimo" (see Figure 3). Adverbs related to other Effort factors can be similarly applied.

In the third series of animations we pay more attention to using adverbs that more likely have an impact on torso shaping. For instance, "walk sadly (discouragedly)" results in a concave shape and slouched shoulders to reveal the inner turmoil; "walk proudly (boastfully)" makes the character strut with chest out and head high. These examples demonstrate the important role that the torso plays in capturing the subtleties of movements. Our belief, also reflected in these examples, is that movements localized in the limbs alone lack conviction and naturalness, if the rest of the body is not appropriately engaged.

8 Conclusions

We have introduced a new paradigm in which a user can produce a wide range of expressive, natural-looking movements by specifying manners and attitudes using natural language verbs and adverbs. While skilled animators can create life-like, expressive characters by controlling the individual lowest level degrees-of-freedom via keyframing or inverse kinematics methods, our approach provides a more systematic and meaningful way of describing qualitative aspects of movements with intuitive natural language instructions. More importantly, by allowing users to customize basic movements with general, qualitative, natural language terms to support individualized expressions, this paradigm may eventually lead to the automatic generation of expressive movements from speech text, a storyboard script, or a behavioral simulation.

9 Acknowledgments

We would like to thank the many students, staff, and colleagues in the Center for Human Modeling and Simulation who make this effort possible. In particular, special thanks go to Diane Chi, William Schuler, Christian Vogler, Koji Ashida, Seung-joo Lee, Fred Azar, Rama Bindiganavale, and Ying Liu, Karen Carter for contributing their time, effort and expertise.

This research is partially supported by U.S. Air Force F41624-97-D-5002, Office of Naval Research K-5-55043/3916-1552793, DURIP N0001497-1-0396, and AASERTs N00014-97-1-0603 and N0014-97-1-0605, DARPA SB-MDA-97-2951001, NSF IRI95-04372, SBR-8900230, and IIS-9900297, Army Research Office ASSERT DAA 655-98-10471, NASA NRA NAG 5-3990, and Engineering Animation Inc. Thanks MIRALab for allowing us to use their face models for research. Support for Monica Costa by the National Scientific and Technological Development Council (CNPq) of Brazil is also gratefully acknowledged.

References


