A Parameterized Schema for Representing Complex Gesture Forms

Huaguang Song and Michael Neff

University of California, Davis, One Shields Avenue, Davis, CA 95616, USA,
{hso, mpneff}@ucdavis.edu

Abstract Gestures can take on complex forms that convey both pragmatic and expressive information. When creating virtual agents, it is necessary to make fine grained manipulations of these forms to precisely adjust the gesture's meaning to reflect the communicative content an agent is trying to deliver, character mood and spatial arrangement of the characters and objects. This paper describes a gesture schema that affords the required, rich description of gesture form. Novel features include the representation of multiphase gestures consisting of several segments, repetitions of gesture form, a map of referential locations and a rich set of spatial and orientation constraints. In our prototype implementation, gestures are generated from this representation by editing and combining small snippets of motion captured data to meet the specification. This allows a very diverse set of gestures to be generated from a small set of input data. Gestures can be refined by simply adjusting the parameters of the schema.

Keywords: Virtual Agent, Animation Frameworks, Behavior Planning, Composite Gesture Representation, Language, Generation

1 Introduction

Hand gestures are one of the most important components of non-verbal communication, conveying both affective and pragmatic information. They involve the hand's path through space, its orientation, and its shape over time; the latter defined by the angles of the finger joints. Being able to rapidly and precisely generate and edit gestures – and adapt them to the current communicative context – is crucial.

Although the motion capture process can be done again and again to try to get all possible data needed to perform certain tasks, it is time-consuming, not cost-effective and will fail where the full context of a character's interactions cannot be predicted ahead of time (i.e. in most interactive scenarios). Developing a framework that allows character systems to reuse mocap data to generate a communicative gesture is essential. A growing set of animation tools [4,6] rely on gesture animation libraries. It is important to be able to generate large numbers of such animations quickly and at low cost in order to build libraries for particular characters and tasks.

Many gestures have quite complex form. Consider a gesture to accompany the text “You mark here, here, here and here and then draw back." that points to a series of locations and then makes a pull back motion to illustrate setting up a carpentry operation. The gesture can naturally be decomposed into a series of segments, one corresponding to each point and the pull back. If motion capture is used, new data would be needed any time the number of marks, their spacing, or the timing of the text changed. With the proper representation, we should be able to adapt such data to any of these changes. The schema presented here solves this problem.

As another motivating example, consider the manner in which people set up a referential space in front of them, locating particular ideas in particular locations. Gestures must be adapted to reflect this organization, requiring a schema that can accommodate location constraints. In other cases, it is possible to build complex gestures out of simple forms when the team did not have the foresight to anticipate a particular gesture need during a motion capture session.

This paper develops a schema for complex gesture forms, describes how the schema can be built from data with minimal user effort and shows an example implementation using this schema. In this work, complex gestures are composed of multiple segments, generally taken from a motion database, and the schema allows fine-grained, algorithmic control of the gesture form. Editing an existing schema allows new gestures to be created and these edits can either be hand specified or generated by the character control system in response to the current context.

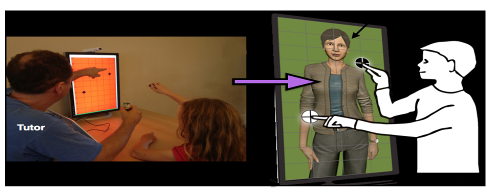

Figure 1. Multimodal tutorial tactics for mathematics. Left: a human tutor teaches proportion. Right: a virtual agent performs the same task

One application of this parameterized schema will incorporate with our collaborative research project (see Figure 1), where we try to replace a human tutor with an intelligent virtual agent to teach children to learn mathematics proportion on multitouch devices. As we analyze the human tutor's behavior, we know the human tutor will respond to children with a variety of interactive gesture forms to guide children to solve the problem. Since the virtual agent will perform the same task as the human tutor, these responses must be composed and performed in real time (i.e. move the hands up and down to specific location). Our representation is developed to meet this specific goal, but it is also fit in other area of where fine-grain adjustment of hand gestures is needed.To simplify the overall process, our system also support semiautomatic segmentation, which allows user mark segments position a full annotate gesture dataset to generate a segmented gesture database based on user interest. Mean while, it will automatically output the corresponding representation schema, this allows user can quick edit parameters rather then build the schema from scratch.

2 Related Work

Badler et al.'s pioneering work on the Parameterized Action Representation (PAR) sought to allow virtual agents to be controlled with natural language instructions [3,2,1]. It introduced a parameterized representation for actions that was based both on affordances of humanoid agents and the language constructs used to instruct them. We also develop a parameterized representation for motion, but with a more narrow focus on representing complex gesture forms rather than a general representation of actions and spatial relationships.

Kopp et al. [5] introduced the Behavior Markup Language, or BML, a high level motion behavior representation for virtual agents. Although we have a format similar to BML and and its derivatives, these language do not support fine-grained gesture editing. They are more focused on simple gesture forms and the coordination between gesture and speech.

3 Generating Schema

While schema can be authored by hand, the more common workflow is to build them from existing motion capture data. In our system, users can load motion capture data of existing complex gesture forms and interactively segment them by scrubbing through a full skeletal playback of the motion, along with gesture trails. They can click to indicate segment divisions at any point in the motion and can specify separate segmentation for the right and left hand. The system will then write out a completely formed schema that can be used to generate the given gesture, along with all the required data. Creating a family of related gestures then only requires small edits to this schema, for instance changing the number of repetitions, changing spatial constraints or changing the path of one of the components. This allows a single motion captured form to be repurposed to generate a large set of gestures that are adapted to the current physical and conversational context of the character.

Table 1 shows the complete representation of a composite gesture.

Figure 2 shows a system architecture for generating motion from our schema.

Acknowledgments

This research was supported by NSF grant IIS 1320029. We thank our collaborators Dor Abrahamson, Seth Corrigan and Virginia J. Flood, who provided insight and expertise that greatly assisted the research.

Table 1. Complete Representation

STARTTIME, ENDTIME, HANDTOUSE, GESTURESEGMENTS, [OVERALLCONSTRAINT], CONNECTIONTYPE
GESTURESEGMENTS = **segment**$^+$
SEGMENT = **Path**, [**HandMotion**], [**Head**]
PATH = **Duration**, **Hold**, (**Data**|**PathType**), (**Touch**|**Point**), [**StartPos**], **Repeat**
PATHTYPE = (“**Straight**″|“**Cyclic**″|“**Curved**″)
STRAIGHT = **EndLoc**
CURVED = **EndLoc**, **Curvature**, **Longitude**
CYCLIC = (“**Clockwise**″|“**Counterclockwise**″), **Curved**
CURVATURE = (“**1LowCurve**″|“**2MidCurve**″|“**3HighCurve**″)
TOUCH|POINT = **Space**, [**Orientation**], [**Duration**]]
SPACE = [**StartOffset**], **Frame**, **CoordinateLoc**, **Tolerance**, [**ApplicationZone**]
FRAME = (“**World**″|“**Chest**″|“**Head**″)
APPLICATIONZONE = (“**Line**″|“**Cube**″|“**Sphere**″|“**BodyLoc**″)
LINE = (“**Coordinate**″|**Coordinate**$^+$)
BODYLOC = (“**Head**″|“**Nose**″|“**Lips**″|“**Shoulder**″|“**Forearm**″|...**etc**)
ORIENTATION = **StartOffset**, **Duration**, **StartPose**, **EndPose**
HANDMOTION = **StartOffset**, **Duration**, **Hold**, **Fingers**
FINGERS = **StartOffset**, **Duration**, **T**, **I**, **M**, **R**, **L**, (**FingerMotion**|**FingerPose**), **Repeat**
T|I|M|R|L = (“**True**”|“**False**”)
HEAD = **StartOffset**, **Duration**, **LookAtLoc**

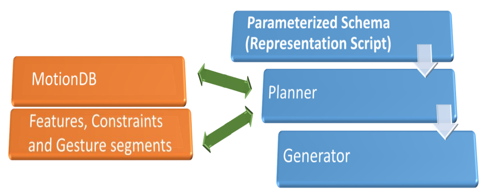

Figure 2. Parameterized Representation system Architecture

References

1. J. Allbeck and N. Badler. Representing and parameterizing agent behaviors. In H. Prendinger and M. Ishizuka, editors, *Life-Like Characters*, Cognitive Technologies, pages 19–38. Springer Berlin Heidelberg, 2004.
2. N. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao, and M. Palmer. Parameterized action representation for virtual human agents. In *Embodied Conversational Agents*, pages 256–284. MIT Press, 2000.
3. N. Badler, R. Bindiganavale, J. Bourne, M. Palmer, J. Shi, and W. Schuler. A parameterized action representation for virtual human agents. *Workshop on Embodied Conversational Characters*, 1998.
4. A. Heloir and M. Kipp. Embr–a realtime animation engine for interactive embodied agents. In *Intelligent Virtual Agents*, pages 393–404. Springer, 2009.
5. S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. Vilhjálmsson. Towards a common framework for multimodal generation: The behavior markup language. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents*, IVA'06, pages 205–217, Berlin, Heidelberg, 2006. Springer-Verlag.
6. M. Neff, M. Kipp, I. Albrecht, and H.-P. Seidel. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions on Graphics (TOG)*, 27(1):5, 2008.