To Gesture or Not to Gesture: What is the Question?

Norman Badler
Monica Costa
Liwei Zhao
Diane Chi
Center for Human Modeling and Simulation
University of Pennsylvania
Philadelphia, PA 19104-6389
215-898-5862
badler@central.cis.upenn.edu

Abstract

Computer synthesized characters are expected to make appropriate face, limb, and body gestures during communicative acts. We focus on non-facial movements and try to elucidate what is intended with the notions of "gesture" and "naturalness". We argue that looking only at the psychological notion of gesture and gesture type is insufficient to capture movement qualities needed by an animated character. Movement observation science, specifically Laban Movement Analysis and its Effort-Shape dimensions with motion phrasing provide essential gesture components. We assert that the expression of movement qualities from the Effort dimensions are needed to make a gesture naturally crystallize out of abstract movements. Finally, we point out that non-facial gestures must involve the rest of the body to appear natural and convincing. A system called EMOTE has been implemented which applies parameterized Effort and Shape qualities to movements and thereby forms improved synthetic gestures.

1 The Problem of Gestural Movements

People move their bodies for many reasons. Some movements are voluntary, such as doing tasks, walking to get somewhere, or speaking. Other movements are involuntary and occur for physical or biological purposes, such as blinking, balancing, and breathing. But a wide class of movements falls inbetween these two; in general this class may be characterized as consisting of movements which occur in concert and perhaps unconsciously with other activities. We can in turn divide this class into at least two sub-classes. One consists of low-level motor controls that assist the accomplishment of a larger coordinated task: thus (unconscious) finger controls make grasps, leg coordination makes walking or running, and lip movements make speaking. Another interesting sub-class is the set of movements which correlate with communicative acts: facial expressions, limb gestures, and body posture. While computer animation researchers have actively studied all these classes of human movements, it remains difficult to procedurally generate convincing ("natural") movements within this last class. Our goal is to explore this problem as it applies to non-facial movements and suggest an approach to a solution.

Let us re-state the problem this way: What parameters characterize body or limb motions in real people performing communicative acts? The primary computational approach to this issue has been through the gesture models proposed by McNeil [13] and elaborated with computer implementations primarily by groups led by Cassell [7, 15, 6], Badler [7, 1], and Thalmann [4, 5]. The primary tenet of the McNeil approach is to characterize communicative arm gestures into several categories:

- *Iconics* represent some feature of the subject matter, such as the shape or spatial extent of an object.
- *Metaphorics* represent an abstract feature of the subject matter, such as exchange, emergence, or use.
- *Deictics* indicate a point in space that may refer to people or spatializable things.
- *Beats* are hand movements that occur with accented spoken words and speaker turn-taking.
- *Emblems* are stereotypical patterns with understood semantics, such as a goodbye wave, the OK-sign, or thumbs-up.

Such an approach has served to make conversational characters appear to gesture more-or-less appropriately while they speak and interact with each other or actual people. The impression that one gets when watching even the most recent efforts in making convincing conversational characters is that the synthetic movements still lack some qualities that make them look "right." Indeed, the characters seem to be doing the right things, but with a kind of child-like and amateurish awkwardness that quickly tags the performance as synthetic. It is not a computer animation problem *per se* – conventional but skilled key-pose animators are able to produce excellent 3D characters. So there is some gap between what such an animator intuits in a character (and is therefore able to animate) and what happens in a procedurally synthesized movement. Key pose animators have managed to bridge the technology gap by careful application of classic rules for conventional animation [14, 12].

The McNeil/Cassell approach to gesture is rooted in psychology and experimental procedures which use human observers to manually note and characterize a subject's gestures during a story-telling or conversational situation. The difficulty in this approach is hidden within the decision to call something a gesture or not. That is, the observer notes the occurrence of a gesture and then records its type. What is not recorded is whether or not the gesture was made: the parameters of movement that made this gesture appear and not another, or what made this gesture appear at all. This issue is crucial in the studies of Kendon [10], who tries to understand the deeper question: What makes a movement a gesture or not? In his work, a gesture is a particular act that appears in the arms or body during discourse. There may be movements which are not gestures and there may be movements which are perceived as gestures in some cultures but not in others. So clearly, the notion of "gesture" as a driver for computer-generated characters cannot be – in itself – the primary motivator of naturalness.

2 The EMOTE Approach

To address this apparent dilemma, we first look toward movement representations outside the constraints of communicative acts. We have been looking at human movement notation systems for decades [2] and have recently seen a breakthrough in building computational models of a particularly important system called Laban Movement Analysis (LMA). The component of LMA that we now study is called Effort-Shape [3]. In her PhD dissertation, Chi created and implemented a kinematic analog to the Effort part of this movement notional system [8, 1]. In recent work we have extended Chi's implementation of arm Effort gestures to include the Shape terms, plus legs and torso [9]. We call this system EMOTE.

LMA has four major components – Body, Space, Shape, and Effort. The EMOTE software covers the Effort and Shape components. Effort describes the qualitative aspects of movement in four motion factors: Space, Weight, Time, and Flow. Each motion factor ranges between *indulging* in and *fighting* against the quality. These extremes 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.

Effort	**Indulging**	**Fighting**
Space	Indirect	Direct
Weight	Light	Strong
Time	Sustained	Sudden
Flow	Free	Bound

Table 1. Effort Elements.

The Shape components in LMA are vertical (rising, sinking), lateral (widening, narrowing), sagittal (advancing, retreating), and flow (outward, inward). In general, shape changes occur in affinities with corresponding Efforts (Table 2 [3]). Although EMOTE allows independent control of shape components, the affinities should normally be respected.

Dimension	Direction	Shape	Effort
vertical	up	rising	Weight-Light
vertical	down	sinking	Weight-Strong
lateral	across	narrowing	Space-Direct
lateral	out	widening	Space-Indirect
sagittal	backward	withdrawing	Time-Sudden
sagittal	forward	advancing	Time-Sustained

Table 2. Shape and Effort Affinities.

The EMOTE system has three features which are critical for resolving our dilemma:

1. A given movement may have Effort and Shape parameters applied to it independently of its geometrical definition in time and space.
2. A movement's Effort and Shape parameters may be varied along distinct numerical scales.
3. The Effort and Shape parameters may be phrased (coordinated) across a set of movements.

In this scheme, the underlying movements are formed by an external process; for example, key time and pose information, a specific gesture stored in a motion library, a procedurally generated motion, or motion captured from live performance. Given such an underlying motion, the EMOTE parameters are applied to it (property 1) to vary its performance. By property (2) we can make the movement more or less distinct in any or all of the Effort-Shape dimensions. By property (3) we can apply property (2) across any series of underlying motions.

Let's consider some examples at this point. Suppose the underlying motions consist of arm movements portraying a single beat gesture that would accompany an accented speech utterance. By adjusting the Effort parameters I can slow down the time course (sustained) and make it more indirect making the gesture change to a vague hand wave. I could start with a slow forward pointing motion and crank it up in the sudden and direct parameters to focus and access the movement into a distinct gesture ("yes, I mean YOU"). By making it rise in Shape, making it more bound, and repeating it (with phrasing to blend together the several occurrences), I can make an agitated or threatening gesture. By neutralizing the Efforts, the gesture seems to almost disappear, becoming a vague forward and upward movement with no emotional overtones and no focus. (Of course, these examples need to be animated and demonstrated on the actual character. We will do so in the accompanying video presentation.)

Our approach to gesture can now been seen as an augmentation of the McNeil/Cassell approach in a missing dimension: gestures of any type exist not just because they have underlying movements but also because they have some distinctiveness in their Effort parameters. This view meshes perfectly with the opposite perspective offered by the LMA proponents: "Gesture ... is any movement of any body part in which Effort or shape elements or combinations can be observed." [3]. Contrapositively, movements lacking expression of these Effort-shape elements may not be considered as a gesture at all. In computational terms, some given underlying movements are modified by EMOTE parameters to express or crystallize the qualities that make it into a gesture.

Actually, we need the rest of the body, not just arms, to make an engaging character. Lamb [11] has observed that a gesture localized in the limb alone lacks impact, but when its Effort-shape characteristics spread to the whole body, a person appears to project full involvement, conviction, and sincerity. By phrasing EMOTE parameters across movements and body parts, perhaps our procedurally animated characters can finally have conviction, too. In the animated Gilbert and George characters produced for [7], our animation technology precluded torso involvement. The characters appear to nod and move their arms in a vaguely disturbing, disembodied fashion. When the rest of the body is forced to move along with the limb gestures, the greater weight of the torso naturally reacts with and absorbs the impact of physical performance.

3 Discussion

We can use this information to analyze why computer generated gesturing characters appear less than natural:

- A character's gestures ought to demonstrate Efforts, otherwise the supposed gestures will look like disinterested movements without the benefit of inner drives, motivators, or personality to back them up.
- A character's gestures should be phrased similarly to communicative phrasing with an expressive content consonant with the principal utterance; for example, a strong accent in speech should be correlated by a strong Effort in gesture. Using a gesture with the same Effort qualities while speaking calmly or excitedly is clearly inappropriate in performance; the Efforts must match or the character will appear to be confused, conflicted, or faked.
- A character moving its arms with appropriate gestures will lack conviction (believability!) if the rest of the body is not appropriately engaged.

This discussion has attempted to blend the psychology of gestures with the structure of movement understanding built from movement observation. By applying the synthesis to a computational model of Effort-Shape, we hope to close the gap between characters created by manual techniques and characters animated by procedures. If the tenets of the movement science behind these observations hold up when transformed into computer code implementations, we should be able to animate engaging, committed, expressive, and – yes – even believable characters consistently and automatically.

4 Acknowledgments

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 655-981-0147, NASA NRA NAG 5-3990, and Engineering Animation Inc. Support for Monica Costa by the National Research and Technology Development Council (CNPq) of Brazil is gratefully acknowledged.

References

[1] N. Badler, D. Chi, and S. Chopra. Virtual human animation based on movement observation and cognitive behavior models. In *Computer Animation Conf.*, Geneva, Switzerland, May 1999. IEEE Computer Society.

[2] N. Badler and S. Smoliar. Digital representations of human movement. *ACM Computing Surveys*, 11(1):19–38, March 1979.

[3] I. Bartenieff and D. Lewis. *Body Movement: Coping with the Environment*. Gordon and Breach Science Publishers, New York, 1980.

[4] P. Becheiraz and D. Thalmann. A behavioral animation system for autonomous actors. In *Workshop on Embodied Conversational Characters*, October 1998.

[5] T. Capin, I. Pandzic, N. Magnenat-Thalmann, and D. Thalmann. *Avatars in Networked Virtual Environments*. Wiley, Chichester, England, 1999.

[6] J. Cassell. Not just another pretty face: Embodied conversational interface agents. *Comm. of the ACM*, 2000. (to appear).

[7] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, W. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics, Annual Conf. Series*, pages 413–420. ACM, 1994.

[8] D. Chi. *A Motion Control Scheme for Animating Expressive Arm Movements*. PhD thesis, University of Pennsylvania, 1999.

[9] M. Costa, L. Zhao, D. Chi, and N. Badler. The EMOTE model for effort-shape. In preparation.

[10] A. Kendon. Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key, editor, *The Relation between Verbal and Nonverbal Communication*, pages 207–227. Mouton, 1980.

[11] W. Lamb. *Posture and Gesture: An Introduction to the Study of Physical Behavior*. Duckworth & Co., London, 1965.

[12] John Lasseter. Principles of traditional animation applied to 3D computer animation. In Maureen C. Stone, editor, *Computer Graphics (SIGGRAPH '87 Proceedings)*, volume 21, pages 35–44, July 1987.

[13] D. McNeil. *Hand and Mind: What Gestures Reveal About Thought.* University of Chicago, 1992.

[14] F. Thomas and O. Johnston. *Illusion of Life: Disney Animation.* Hyperion, New York, 1995.

[15] H. H. Vilhjalmsson and J. Cassell. Bodychat: Autonomous communicative behaviors in avatars. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 269–277. ACM, May 1998.