

Gesticulation Behaviors for Virtual Humans

Liwei Zhao and Norman I. Badler
Center for Human Modeling and Simulation
Department of Computer and Information Science
University of Pennsylvania
Philadelphia PA 19104-6389 USA
lwzhao@graphics.cis.upenn.edu, badler@central.cis.upenn.edu
1-215-898-5862 phone; 1-215-573-7453 fax

Abstract

Gesture and speech are two very important behaviors for virtual humans. They are not isolated from each other but generally employed simultaneously in the service of the same intention. An underlying PaT-Net parallel finite-state machine may be used to coordinate them both. Gesture selection is not arbitrary. Typical movements correlated with specific textual elements are used to select and produce gesticulation online. This enhances the expressiveness of speaking virtual humans.

Keywords: Virtual Human, Agent, Avatar, Gesture, Posture, PaT-Nets

1 Introduction

The past few years have seen several research efforts on human gestures (i.e. [1, 2, 5, 8, 27, 18, 16, 13]). Many of these projects have focused on interpreting human gestures for interactive control. Creating appropriate gestures in a virtual human has not been as well studied because the range of gestures performed during speech output is much larger than a symbolic selection set used for discrete inputs. For example, in [8] four gesture types are distinguished:

- *Iconics* represent some concrete feature of the accompanying speech, such as an object's shape.
- *Metaphorics* represent an abstract feature concurrently spoken about.
- *Deictics* indicate a point in space, and may refer to persons, places and other spatializeable discourse entities.

- *Beats* are small formless waves of the hand that occur with heavily emphasized words, occasions of turning over the floor to another speaker, and other kinds of special linguistic work.

While Cassell's system implemented instances of each type of gesture, the most prevalent were iconics linked to mentions of specific objects, metaphorics linked to specific actions, and beats linked to speech intonation.

Following Cassell's lead, new problems in gesture generation were exposed.

1. Coarticulation: Generating a smooth transition from one gesture to the next without returning to a specific rest pose.
2. Expression: Modifying the performance of a gesture to reflect the agent's manner or personality.
3. Spatialization: Integrating a deictic gesture into the surrounding context.
4. Selection: Generating a metaphoric that might be associated with an abstract concept.

Problem 1, *coarticulation*, has been addressed by a number of computer graphics researchers ([8, 12, 28, 26]), although the issue has other aspects (such as preparatory actions) which remain unsolved. Problem 2, *expression*, is being investigated at a number of places ([6, 33, 10]). In this paper, we investigate problems 3 and 4. Of these two, *spatialization* is easier, since the desired gesture is combined or composited with inverse kinematics to point or align the gesturing body part with the spatial referent. *Selection* entails determining gestures that people would likely interpret and accept as "natural" and "representative." These concepts are orthogonal: a naturally performed (motion captured) gesture might not be appropriate to the speech text, while a synthesized (less natural) arm

motion might nevertheless be representative of the expressed concepts.

The selection problem itself splits into two: one is the creation of the gestural motion and the other is the mapping from the textual content to the gesture. For example, to create a character waving hello during a greeting, one has to create the waving motion as well as know when to invoke it upon encountering a greeting context. In this work we assume that the motions themselves are generated by inverse kinematics, motion capture, or otherwise pre-created (e.g. key pose) sequences. Our contribution lies in proposing a representative mapping from concepts to gestures such that they are selected based on stylized rhetorical speaking.

To select and spatialize various gestures correlating speech and language, we use an underlying coordination scheme called **PaT-Nets** [3]. The virtual human animation is implemented as an extensions to *Jack*¹. The inputs (see below) to the system are in the form of speech texts with embedded commands, most of which are related to gestures. The gestures are controlled by **PaT-Nets** to coincide with the utterance of the speech. While the embedded commands in our examples are manually inserted for now, the idea is to detect the presence of the corresponding concepts in the raw text stream and automatically insert the deictics and metaphoric based solely on the words used.

```

Hello, \gest_warning welcome.
\head_front Currently, I can support following basic arm gestures.
Now let me introduce you some simple objects I know:
\point_idxf{table.table.corner} this is a table
\point_idxf{door.door.panel} this is a door
\point_idxf{chair1.chair.red} this is red chair
\point_idxf{chair0.chair.yellow} this is yellow chair
\head_slant_right Let me show you the basic arm gestures
\gest_arm_reject arm reject gesture
\gest_arm_unlikely arm unlikely gesture
\gest_arm_not arm not gesture
\gest_arm_improbable arm improbable gesture
\gest_arm_doubtful arm doubtful gesture
\gest_arm_probable arm probable gesture
\gest_arm_is arm it is gesture
\gest_arm_certain arm certain gesture
\gest_arm_obvious arm obvious gesture
\gest_arm_enchanting arm enchanting gesture
\gest_arm_absolute arm absolute gesture
Next, let me show you some hand gestures:
\hand_convulsive{plane0.plane.stand1} convulsive hand gesture
\hand_expanded{plane0.plane.stand1} expanded hand gesture
\hand_exasperation{plane0.plane.stand1} exasperation hand gesture
\hand_authority{plane0.plane.stand1} authority hand gesture
\hand_relaxed{plane0.plane.stand1} relaxed hand gesture
\hand_exalation{plane0.plane.stand1} exalation hand gesture
\hand_conflict{plane0.plane.stand1} conflict hand gesture
\hand_prostration{plane0.plane.stand1} prostration hand gesture
\head_slant_left
Finally, I can support following basic general gestures:
reject gesture \gest_reject
give and take gesture \gest_givetake
warning gesture \gest_warning

good bye

```

2 Gesticulation

An agent or avatar may have a wide variety of movement behaviors, but we focus our attention on ges-

¹ *Jack* is a software product from Transom Technologies, Inc.

tures and speech. Kendon [20] offers a distinction between *autonomous gestures* (gestures performed without accompanying speech) and *gesticulation* (gesture performed concurrently with phonological utterance). Gestures and speech are closely associated together. They are generally employed simultaneously in the service of the same intention. Well-coordinated gestures and speech enhance the expressiveness and believability of speaking virtual humans. In this paper, we restrict our investigation to *gesticulation*.

2.1 Gestures

The study of gestures in dance and oratory may date back to the beginning of seventeenth century [7]. More recently, semioticists from the fields of anthropology, neurophysiology, neuropsychology and psycholinguistics (Freedman [17]; Wiener, Devoe, Rubinow and Geller [34]; McNeill and Levy [24]) have been interested in the study of gestures. The *Lexis* dictionary (1977) gives the most general definition of gesture — “movements of body parts, particularly the arms, the hands or the head conveying, or not conveying, meaning.”

While gestures are the “little” movements that are confined to a part or parts of the body, if just considered in isolation they have very limited contribution to make to non-verbal communication. (Emblems and manual languages, such as American Sign Language, are exceptions because the communication is fully borne by movements.) Gestures are rarely performed outside a communicative context and only occasionally transmit any depth of emotion or information, since, as soon as there is any complicated meaning, the gestures can only be “read” in relation to the whole expressive movement of the body [14, 9, 22].

Most of the current research in gestures is related to computer vision, human-computer interaction, and pattern recognition, where the gestures are mainly studied in isolation [27, 18, 16, 13]. However, gestures used by an agent or avatar in a virtual environment are quite different. First, it is a process, not a fixed posture. For example, when someone waves a hand, it is not the final position of the hand which is the proper object of study, but the *process by which it got there* — the actual process of movement. Secondly, it is almost always accompanied by other gestures or communicative channels.

In the following we study arm, hand, and head gestures. Above all, we recognize that gesticulation has its limitations. The interpretation might be both culturally oriented and individually biased. Personality and social context may constrict or amplify the motions.

But in general we seek to set a baseline of gesticulatory behavior which can then be parameterized and modified by other means.

2.1.1 Arm Gestures

Human arms serve at least two basic separate functions [1]: they allow an agent/avatar to change the local environment through dextrous movements by reaching for and grasping objects [19, 15]; and serve social interaction functions by augmenting the speech channel with communicative emblems, gestures and beats [8].

A well-performed arm gesture, accompanied by proper hand gestures, plays an important role in integrating some deictic gestures into the surrounding context (*spatialization* problem) and reflecting the agent's manner or personality to some extent (*expression* problem). For example, in [30] it was noted that arm gestures with different inclinations indicate different degrees of affirmation — from 0 (straight down) to 45 degrees indicates neutral, timid, cold; from 45 degrees to 90 degrees, expansive and warm; and from 90 degrees to 180 degrees, enthusiastic (see Figure 1). We implemented this series of stereotypical arm gestures as a representative (metaphorical) mapping from affirmation concepts to gestures such that they can be correlated with the degree of affirmation in a speech.

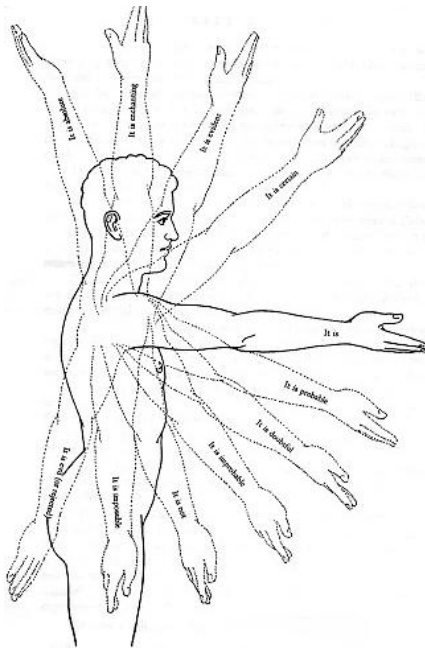


Figure 1: Arm gestures with different degrees indicate different degrees of affirmation (taken from [30])

2.1.2 Hand Gestures

The hand is the most fluent and articulate part of the body, capable of expressing almost infinite meanings. Hand gesture languages have been invented by communicative needs and by the deaf communities of various cultures. The classic gesture languages of the Hindu dance contains about 57,000 cataloged hand positions each having the specific value of a word or explicit and distinct meaning [29]. It is virtually impossible to implement all these hand gestures. In this paper, we investigate the *selection* problem. So we focus on the hand gestures which can easily generate a *metaphoric* associated with an abstract concept. In the system, the virtual human agent attempts to use hand gestures that are more selective and which are much more closely coordinated with what is being said in words. For example, when attempting to offer a definition of a word such as “write,” the agent may pantomime the writing action while vocalizing the verbal definition [8].

Delsarte [30] provided a small set of stereotypical hand gestures correlated with grasping, indicating, pointing, and reaching (illustrated in Figure 2). We implemented all these hand gestures and they can be performed either by left or by right hand, with preference for the right hand under default circumstances. To avoid crossing the arm over the body and to keep the body posture open, the nearer hand to the target object is always used. In addition, every hand gesture is coordinated with head and eye orientation, arm gestures, and vocalization, all of which are employed simultaneously in the service of interpreting an abstract concept.

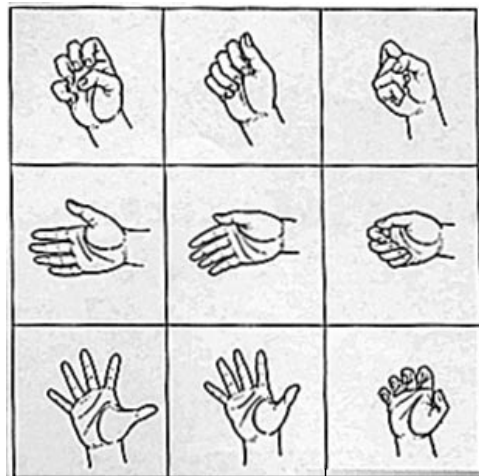


Figure 2: Grasping, indicating and reaching hand gestures (taken from [30])

2.1.3 Head Gestures

The head can be a very effective gesturing tool. The face is one of the most important parts in computer animation. It can be divided into three zones: (1) the forehead and eyes; (2) the nose and upper cheek; (3) the mouth, jaw, and lower cheeks [22]. The eyes in turn have three components — the eyeballs, the eyelids, and the eyebrows. In [23] 405 combinations of these components alone are listed. When these uses are combined with expressions of the mouth, and the attitudes or position of the head — the possible combinations are almost beyond computing. Again, we focus our attention on those head gestures that are related to the spatialization and selection problems.

Different from arm gestures and hand gestures, head gestures are employed more selectively. For example, a 10-year-old gestures elaborately using arms or hands while he is talking as if, as Freedman puts it [17], “he surrounds himself with a visual, perceptual and imagistic aspect of his message.” On the other hand, the head gestures are used very selectively, usually only in relation to specific words, with which the head gestures are highly coordinated.

Delsarte [30] gave 9 positions or attitudes of head gestures combined with eyes (as shown in Figure 3), which we think can be acted as a set of representative of head gestures that help express abstract concepts gesturally.

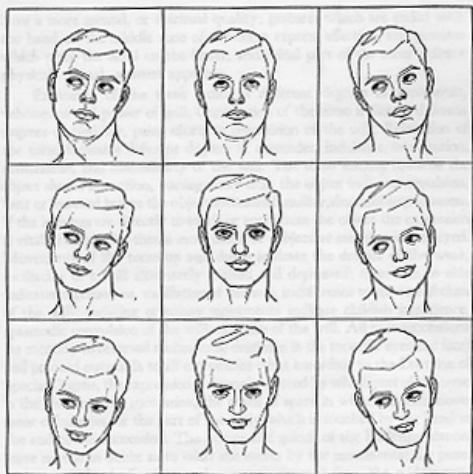


Figure 3: Head gestures combined with eyes (taken from [30])

2.2 Postures

Postures are highly correlated with speech. We usually use the postures as interpretative tools to understand the speech and we don't allow ourselves to be influenced by words which may be quite at variance with what is being “said” in the silent postures. In our gesticulation system, to avoid having the words discounted, a virtual human agent usually adopts a neutral posture — standing up straight with both feet slightly apart and firmly planted on the floor, and should adopt an orientation and eye gaze facing to the audience.

Postures are also highly correlated with gestures. Within a sequence of movements a small gesture, such as waving and smiling, may be very significant, but it is also significant as a part of the whole body. Gestures need postures as a background [22, 9, 14]. On the other hand, postures almost always have gestures going on around them. Together gesturing and posturing make up the process of movement. Postural semantics has received very little systematic attention in virtual human research. Lamb and Watson [22] note that posture is an individual characteristic, and is highly influenced by the conventions of the society. DeWall et al. (1992) provide methods improving sitting postures of CAD/CAM workers. Ankrum (1997) reports the interrelationships between gaze angle and neck posture. Tsukasa Noma (1997) uses posture as a visual aid to presentation. But in none of these efforts is the interdependence between gestures and postures addressed.

In our gesticulation system we implemented two of the postures given by Delsarte: both are related to standing and may be either merged with, or segregated from, various gestures.

2.3 Locomotion

In order to expressively interpret an abstract concept, an agent or avatar might interact with an object which visually corresponds to the concept being interpreted. The interaction includes detecting, orienting to, locating, reaching, and pointing to a visual object. It can be argued, though, that these interactions can be distinguished according to the (spatial) field in which they occur. In fact, these interactions can occur either in immediate surroundings in which reaching, indicating or grasping is achieved without locomotion, or in the visual field outside of direct reaching and grasping.

Therefore, to interact with a target object, an agent or avatar must determine if she is within a suitable distance from the target. Otherwise, she must first walk to an action-dependent position and orientation (pre-action) before the initiation of the specified action. Af-

ter completing the action, she must decide if she needs to walk to the next action-dependent position and orientation (post-action). Also, she must keep in mind an explicit list of objects to be avoided during the locomotion process. Such decision-making and walking are coordinated by **PaT-Nets**.

3 The Underlying Coordination Model

3.1 Coordination via PaT-Nets

Using traditional animation techniques, human behavior is defined as a set of linear sequences which are determined in advance. During motion transitions, a motion generator has to monitor the whole transition from the current motion to the next one [3]. This gives the animator great control over the look and feel of the animation. Anyone who goes to the movies can see marvelous synthetic characters such as aliens, Martians, etc. However, all these characters are created typically for one scene or one movie and are not meant to be re-used [1, 26]. Should the same techniques be used in virtual humans, it would greatly limit their *autonomy, individuality*, and therefore believability.

Some researchers have attempted to get around this problem by breaking the animation down into smaller linear sequences and then switching between them contingent upon user input. So the main concern is dealing with the transitions between these sequences. The simplest approach is to ignore the transition and simply jump from one motion to the next. This works in situations where fast transitions are expected, but appears jerky and unnatural when being applied to virtual humans. Another approach is to have the beginning and ending in the same standard posture, thus eliminating the instantaneous jump. While this approach offers smooth continuous motion, beginning and ending each motion in the same still posture is very unnatural: each time the body needs to return to a “neutral” (generic intermediate) posture before the next motion can begin. Moreover, the transitions between motions need to be defined for every pair of motions in advance. In NYU’s Improv Project [26] they proposed a technique called motion blending to automatically generate smooth transitions between isolated motions without jarring discontinuities or the need to return to a “neutral” pose. But the motion generator still needs to assign joint angles to the whole body. In some sophisticated scenarios where agent and avatars are engaging in some complex behaviors and interactions, this becomes ineffective. Using **PaT-Nets**, groups of body parts are

assigned to individual nets: WalkNet, ArmNet, HandNet, FaceNet, SeeNet and SpeakNet. All these nets are organized in a hierarchical way. (The structure of the nets is shown in Figure 4.) It makes the interaction between agents/avatars [8] and synchronization of movements relatively easy, because the action generator (ParserNet) is not involved in directly assigning joint angles to the whole body: instead it sends messages to designate individual nets to do the job, hence its main function is coordination. For example, to move a hand, ParserNet does not need to directly assign joint angles. All it needs to do is to send a message to the GestureNet, which in turn sends a message to the HandNet. Then the HandNet moves the joints depending on the timing and joint angles in the message. This coordination can be applied to the game of “Hide and Seek” [4], two person animated conversation [8], simulated emergency medical care [11], and TV presenter or weatherman [25].

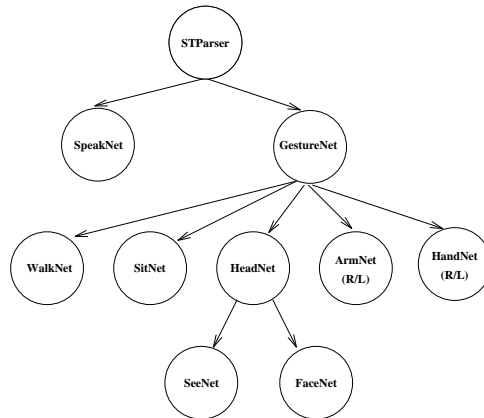


Figure 4: PaT-Nets for gesticulation behaviors

3.2 PaT-Nets

PaT-Nets (Parallel Transition Networks) are finite state machines that can execute motions effectively in parallel. The original **PaT-Nets** were implemented in lisp by Welton Becket [3]. In order to maximize real-time animation control, Tsukasa Noma re-implemented the **PaT-Nets** in C++, with further modifications made by Sonu Chopra. Each class of **PaT-Nets** is defined as a derived class of the base class LWNet, which stands for Light Weight PaT-Nets. They have the following properties:

- Two or more PaT-Nets can be simultaneously active.

- Two or more nodes can be simultaneously active in a PaT-Net. It enables us to represent simple parallel execution of actions in a single PaT-Net.
- PaT-Nets can call for actions and make state transitions either conditionally or probabilistically.
- All active PaT-Nets are maintained on a list called the *LWNetList*. This list is scanned every clock tick.
- *Jack* commands can be invoked within PaT-Nets to manipulate any *Jack* data structure.

Currently **PaT-Nets** support 9 different node types: Normal, Call, PAL, Join, Indy, Kldp, Monitor, Exit and Halt. Normal node is used to execute an action and Call node is used to call a function. The action/call is preceded by a pre-action and succeeded by a post-action. Transition to one of a set of post-actions depends on the action's boolean function or the pointer returned by the call function. All the nodes spawned from PAL node should be done in parallel. Join/Indy/Kldp nodes link the spawned nodes: the differences are that the Join node waits for all spawned nodes to be finished before moving on to the next node; the Indy node moves to the next as soon as the first spawned node is done and leaves the remaining spawned nodes untouched; and Kldp is similar but kills the remaining spawned nodes. The Join/Indy/Kldp nodes make synchronization possible. The Monitor node checks the monitor condition every clock tick and activates the monitor action whenever the condition is evaluated true. The Halt node simply terminates the current PatNet node, but the Exit node removes the current PatNet from the active LWNetList. For example, in the movements shown in Figure 5, the Walk node is first executed. Then the PAL node spawns a Speak node and a series of sequential actions defined by a Gesture node, a Normal node, and a PointAt node. The Speak node should be run simultaneously with the sequential actions. Basically this is walking followed by speech and a pointing gesture in parallel.

4 Results

We implemented the gesticulation system on an SGI Onyx/RealityEngine. In the current implementation, **PaT-Nets** are extended to contain twelve different nets that can be running simultaneously. The motion generator (ParserNet) contains 66 nodes to synchronize different movements: now it can support up to 2 postures, 3 head gestures, 12 arm gestures and 12

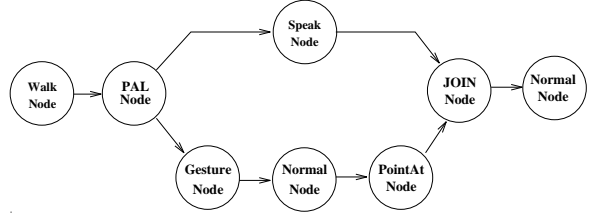


Figure 5: A PaT-Nets example: walking, speaking and pointing

hand gestures. During the animation, the virtual human agent walks around the room and points out some interesting objects such as table, door, red chair, yellow chair, etc. (We do not yet deal with automatically recognizing the objects in the virtual environment; instead as a pre-processing step we associate sites in the coordinate system with the objects.) Then he walks to the front scene and demonstrates some arm gestures and hand gestures (Figures 6 and 7).

Animations are generated in real-time (30 frames per second). For voice output, we use an Entropic Research Laboratory *TrueTalkTM* TTS (Text-To-Speech) system [32] running on an SGI Indigo2. The gesture movements are controlled by **PaT-Nets** to coincide with the utterance of the speech.

5 Conclusions

We discussed a virtual human gesticulation system where typical gestures correlated with speech are used to select and produce gesticulation in real time. We also investigated the *Spatialization* and *Selection* problems and proposed a representative mapping from concepts to gestures such that they are selected based on stylized rhetorical speaking. An underlying coordination mechanism called **PaT-Nets** is employed to select and spatialize various gestures associated with speech and language.

In our current implementation, there is still much work to do for the near future:

- Add more nodes to FaceNet to improve the facial expression and mimic the mouth movements more precisely during speech.
- Add more gestures/movements which are necessary in a dialogue structure, and environment- and object-sensitive interaction.
- Transport all gestures/movements to JackMOO [31] to expand the scope and range of human ac-



Figure 6: Examples: arm gestures by Jack

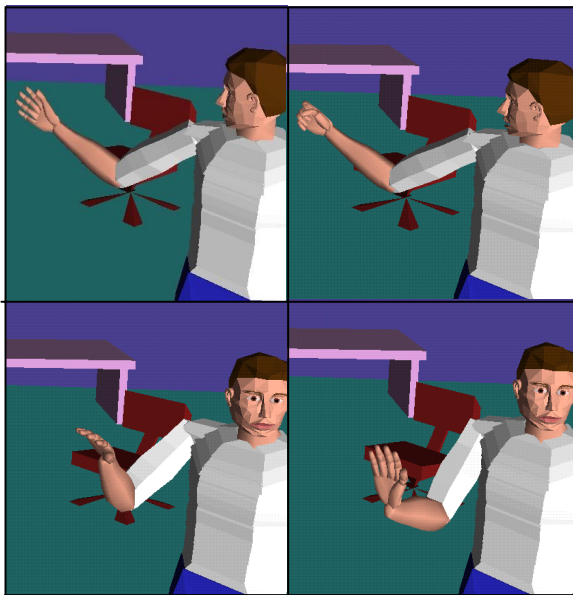


Figure 7: Examples: hand gestures by Jack

tions that an avatar must portray in a web-based virtual environment.

6 Acknowledgments

The authors would like to thank Sonu Chopra, Rama Bindiganavale and Pei-Hwa Ho for discussions and technical support. This research is partially supported by U.S. Air Force through Delivery Orders #8 and #17 on F41624-97-D-5002; Office of Naval Research (through Univ. of Houston) K-5-55043/3916-1552793, DURIP N0001497-1-0396, and AASERTs N00014-97-1-0603 and N0014-97-1-0605; Army Research Lab HRED DAAL01-97-M-0198; DARPA SB-MDA-97-2951001 (through the Franklin Institute); NSF IRI95-04372; NASA NRA NAG 5-3990; National Institute of Standards and Technology 60 NANB6D0149 and 60 NANB7D0058; SERI Korea, and JustSystem Japan.

References

- [1] N. Badler. Real-time Virtual Humans. *Pacific Graphics*, 1997.
- [2] N. Badler. Virtual Humans for Animation, Ergonomics, and Simulation. *IEEE Wrkshp. on Non-Rigid and Articulated Motion*. Puerto Rico, June 1997.
- [3] N. Badler, C. Phillips, and B. Webber. *Simulating Humans: Computer Graphics, Animation, and Control*. Oxford University Press, New York, 1993.
- [4] N. Badler, B. Webber, W. Becket, C. Geib, M. Moore, C. Pelachaud, B. Reich and M. Stone. Planning for animation. *Computer Animation* N.Magenat-Thalmann and D.Thalmann, editors. Prentice-Hall, 1996.
- [5] R. Boulic, P. Becheiraz, L. Emering, and D. Thalmann. Integration of motion control techniques for virtual human and avatar real-time animation. pp. 111-118 Proc. *VRST '97*, ACM Press, 1997.
- [6] A. Bruderlin and L. Williams. Motion signal processing. Proc. SIGGRAPH 1995, pp. 97-104.
- [7] J. Bulwer, *Chirologia: Or the natural language of hand and Chironomia: or the manual art of rhetoric*. Southern Illinois University Press, Carbondale, IL. 1994.
- [8] 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*, pp. 413-420. ACM, 1994
- [9] M. Cranach and I. Vine. *Expressive Movement and Non-verbal Communication*, Academic Press, London, 1975.

- [10] D. Chi. Animating expressivity through effort elements. PhD Dissertation, in progress, University of Pennsylvania, 1998.
- [11] D. Chi, B. Webber, J. Clarke, and N. Badler. Casualty modeling for real-time medical training. *Presence*, 5(4):359-366, 1995.
- [12] M. Cohen and D. Massaro. Modeling coarticulation in synthetic visual speech. In M. Thalmann and D. Thalmann, eds., *Models and Techniques in Computer Animation*, Tokyo, 1993. Springer-Verlag.
- [13] Y. Cui, D. Swets and J. Weng. Learning-based hand sign recognition using SHOSLIF-M. *Wrkshp. on Integration of Gesture in Language and Speech*, '96.
- [14] M. Davis, Understanding body movement, an annotated bibliography. New York, Arno Press, 1972.
- [15] B. Douville, L. Levison, and N. Badler. Task level object grasping for simulated agents. *Presence* 5(4) pp. 416-430, 1996
- [16] R. Foulds and A. Moynahan. Computer Recognition of the gestures of people with disabilities. *Wrkshp. on Integration of Gesture in Language and Speech '96*, Wilmington, DE, USA.
- [17] N. Freedman. Hands, words and mind: on the structuralization of body movements during discourse and the capacity for verbal representation, pp. 110-132. *Communicative Structures and Psychic Structures: A Psychoanalytic Approach*. New York and London: Plenum Press, 1977.
- [18] W. Freedman and C. Weissman. Television control by hand gestures *IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition*, Zurich, June, 1995
- [19] J. Gourret, N. Magnenat-Thalmann, and D. Thalmann, Simulation of object and human skin deformations in a grasping task. *ACM Computer Graphics* 23(3) pp. 21-30,1989
- [20] A. Kendon. Current issues in the study of gesture. *The Biological Foundations of Gestures: Motor and Semiotic aspects*, pp. 23-47. Lawrence Erlbaum Associates, Hillsdale, NJ. 1986.
- [21] A. Kendon. Gesticulation and speech: Two aspects of the process of utterance. In M.R.Key(Ed.), *The relationship of verbal and nonverbal communication* The Hague: Mouton Publishers, 1980.
- [22] W. Lamb and E. Watson. *Body Code: The Meaning in Movement*. Routledge & Kegan Paul Ltd. pp. 85-99. 1979.
- [23] Steele Mackaye. Harmonic Gymnastics and Pantomimic Expression. (edited by Marion Lowell) M. Witmark & Sons, New York, 1963.
- [24] D. McNeill *The conceptual basis of language*. Hillsdale, NJ: Lawrence Erlbaum Assoc, 1996.
- [25] T. Noma and N. Badler. A virtual human presenter. In *IJCAI '97 Workshop on Animated Interface Agents*, Nagoya, Japan, 1997.
- [26] K. Perlin and A. Goldberg. Improv: A System for Scripting Interactive Actors in Virtual Worlds. Proc. SIGGRAPH 1996, pp. 205-216.
- [27] F. Quek. Toward a vision-based hand gesture interface. *Proceedings of the Virtual Reality System Technology Conference*, pp. 17-29, August 23-26, 1994, Singapore.
- [28] C. Rose, B. Guenter, B. Bodenheimer and M. Cohen. Efficient generation of motion transitions using space-time constraints. In ACM Computer Graphics, Annual Conf. Series, pp. 147-154, 1996.
- [29] H. Russell, *The gesture language of the Hindu dance*, New York, B. Blom [1964, c1941]
- [30] T. Shawn. Every little movement – a book about Delsarte. M. Witmark & Sons, 1954
- [31] T.J. Smith, J. Shi, J. Granieri, and N. Badler JackMOO: A web-based system for virtual human simulation, *WebSim* Conference, 1998.
- [32] TrueTalk Programmer's Manual, 1995. Entropic Research Laboratory.
- [33] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. Proc. SIGGRAPH 1995, pp. 91-96.
- [34] M. Wiener, S. Devoe, S. Rubinow and J. Geller. Non-verbal behavior and nonverbal communication. *Psychological Review*, pp. 185-210.