Real time multimodal interaction with animated virtual human.

Li Jin[1]
Zhigang Wen[2]

[1] Harrow School of Computer Science, University of Westminster
[2] School of Computer Science, University of Birmingham

Real Time Multimodal Interaction with Animated Virtual Human

Li Jin
Harrow School of Computer Science
University of Westminster
Harrow, HA1 3TP, UK
Email: Li.Jin02@wmin.ac.uk

Zhigang Wen
School of Computer Science
University of Birmingham
Birmingham, B15 2TT, UK
Email: Z.Wen@cs.bham.ac.uk

Abstract

This paper describes the design and implementation of a real time animation framework in which animated virtual human is capable of performing multimodal interactions with human user. The animation system consists of several functional components, namely perception, behaviours generation, and motion generation. The virtual human agent in the system has a complex underlying geometry structure with multiple degrees of freedom (DOFs). It relies on a virtual perception system to capture information from its environment and respond to human user's commands by a combination of non-verbal behaviours including co-verbal gestures, posture, body motions and simple utterances. A language processing module is incorporated to interpret user's command. In particular, an efficient motion generation method has been developed to combines both motion captured data and parameterized actions generated in real time to produce variations in agent's behaviours depending on its momentary emotional states.

1. Introduction

Today software titles demand 'smarter' participants in the simulated virtual world. For instance, in the entertainment industry like PC games, the Non-Player-Characters (NPCs) controlled by the computer are expected to exhibit convincing behaviours in respond to dynamic change of the environment and human player's activities. Advancements from artificial intelligence, computer animation, and human computer interaction have enabled the development of intelligent 3D embodied character that is capable of interacting with human user in a believable way in real time.

There have been numerous efforts to develop the real time animation systems for lifelike embodied agents during the past decade and significant achievements have been made [1, 2, 3, 4]. However, real time animation creation of realistic virtual human still remains as a challenge. One of the major problems for characters in those real time applications is that the character is likely to exhibit repetitive behaviours by relying on some well-canned predefined motion elements. Although a number of motion retargeting algorithms have been proposed [5,6] they are generally too computationally expensive for real time applications. People gesture when they communicate with others. Researchers believe that gesture arises from the same generative process that produces speech [7] and therefore it is closely related to speech. Therefore, in order to create effective interaction between virtual animated agent and human user, the generation and coordination of verbal and non-verbal behaviour play an important 33role in the animation generation process. However, due to the lack of thorough understanding of human mental process for generating speech accompanying gestures, it is difficult to design such an effective computational model although progresses have been made in recent years. For instance, Cassell et al. [8] use linguistic and contextual information contained in the text to control the movements of the hands, arms and face, and the intonation of the voice. The mapping from text to facial, intonational and body gestures was contained in a set of rules derived from the state of the art in nonverbal conversational behaviour research. Kopp [9] described an approach to generate multimodal utterances for the given communicative goals by analyzing the information presented by iconic gesture into semantic units that were linked to hand shape, gesture trajectory. Balder et al. [10] argued that by looking only at the psychological notion of gesture and gesture type is insufficient to capture movement qualities needed by an animated character. EMOTE (expressive motion engine) uses inverse kinematics to control the qualitative aspects of end-effector specified movements. The Effort and Shape model originated from Laban Movement Analysis is used to adjust the character's movement.

Therefore, the main objective of this paper is to design a virtual human animation system with real time performance on a modern PC platform with assistance from 3D graphics hardware acceleration features that have been available to common users. Particularly, a hybrid motion generation method has been developed and incorporated into the animation framework to effectively generating variations in agent's behaviours depending on

Proceedings of the Information Visualization (IV'06)
0-7695-2602-0/06 $20.00 © 2006 IEEE

its emotional states. It worth pointing out that this paper concentrates on the method to produce character's body motions and their variation in real time instead of generating facial expression and other behaviours such as eye gazing, although these behaviours also play important roles in the interaction process.

Section 2 describes the design of the animation system and its major functional components. Section 3 briefly illustrates the implementation and a simple animation example. Section 4 finally draws conclusions.

2. The Animation System

The animation system is divided into 6 main components, namely environments, perception, language processing module for user command, behaviours generation, motion generator and rendering module. Figure 1 shows high-level functional architecture of the framework.

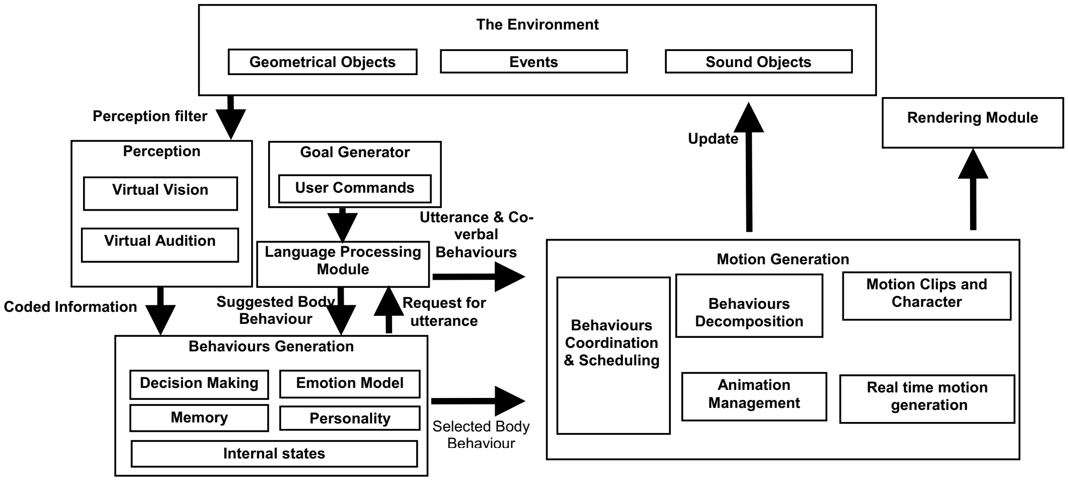


Figure 1. Architecture of animation framework

The *environment* normally contains geometrical objects, sound objects, and events. Geometrical objects refer to 3D objects with vertices and texture maps, which are detectable by the agent's virtual vision. In a complex virtual environment containing large number of 3D objects, spatial partition techniques are useful to arrange these objects in some kinds of hierarchy that accelerate the agent's objects detecting process [11]. Geometrical objects are normally "seen" by the agent with additional properties being memorized such as the location of the object, time of being detected, ID of the object etc. Sound objects in the environment can be detected by the agent's virtual audition sensor, subsequently affecting the agent's behaviours. The environment contains another special object called events. Events can generally be divided into two categories, namely functional events and dynamics events. Functional events can be considered as the "built-in" behaviours of the objects and they can be triggered by under some conditions. For instance, a door in a scene can be opened or closed by the agent. Dynamics event is generally generated in related to agent's activities. This type of event will have significant impact on the agent's emotional state, which will subsequently affect the agent's behaviours or the way the agent execute its behaviours. For instance, the result of an agent's attempt to capture some objects will cause agent become happier or angrier. One of the novelties of the proposed animation system is that it has the mechanism to visualize the subtle change of agent's emotional states via its animated behaviours.

The *Behaviours Generation* contains several key functional components to perform action selection for the agent, namely decision-making, emotion, personality, memory, and internal states. The *Decision Making* component receives coded information from the perception channel and body behaviours suggestion from language processing module which performs a series of linguistics processing of the user command. The *decision making* component relies on a hierarchical action network to perform the action selection for agent according to its perceptual information, internal states, memory and goals. Suggested body behaviours from *language processing* module may be rejected or accepted in this action selection process. Emotion plays an important role in creating believable agent behaviours [12, 13]. Psychological and neuroscience research indicates that emotions have a significant impact on human behaviours, both through their use as a non-verbal communication channel such as gesture, posture, facial expression and so on [14]. It is therefore important to incorporate emotion into our animation system. Emotion model in the system is based on the OCC model [15]. The OCC model specifies how events, agents and objects are appraised according to respectively their desirability, praiseworthiness and appealingness, which are defined by a set of parameters such as goals and attitudes. The process of integrating OCC model into agent behaviours can be divided into 4 steps, namely classification, quantification, interaction, and mapping [16].

It is believed that personality and emotion have significant influences on behaviours and how behaviours are expressed [17,18] Different personality model has been studied in the psychology community such as the OCEAN model [19]. This model has five dimensions, namely openness, conscientiousness, extraversion, agreeableness, and neuroticism. The link between personality and OCC emotion model is described in [20]. Their idea is essentially to construct a *Personality-Emotion Influence Matrix*, in which indicates how each personality factor influences each emotion. A simple emotion update equation as proposed in [20] is therefore used in the animation system.

$$e_{t+1} = e_t + \alpha(p, \omega_t, a) + \beta(p, \omega_t)$$

Function $\alpha(p, \omega_t, a)$ is to calculate the changes of the emotional state based on personality p, emotional state history ω_t, and emotion influence a from OCC model. $\beta(p, \omega_t)$ is calculate the decay of emotional sates based on personality and emotional state history. Emotion in the developed animation system is primarily used to decide how the selected behaviours will be executed depending on agent's momentary emotional states.

The language module is responsible for annotating input text from the user with linguistic information that is later used for behaviours generation module. An English syntax parser named after Link Grammar [21] is used. The Link Grammar Parser is a syntactic parser of English based on link grammar, an original theory of English syntax. Given a sentence, the system assigns to it a syntactic structure, which consists of a set of labelled links connecting pairs of words. The parser also produces a "constituent" representation of a sentence. For instance, the command, "Agent A, go and find the white ball on the grass land quickly", has the following syntactical structure output:

```
(S Agent A ,
   (S (VP go and find
          (NP the white ball)
          (PP on
              (NP the grass land))
    (ADVP quickly))))
```

The language module then encodes such syntactical analysis information into a simple semantics level form as *Find (Agent A, Ball, on the grass land, quickly)*, which is used to suggest an action for the virtual agent to take. The suggested action from language module together with information from the virtual vision and audition channel are fed into the central decision making component for the final action selection. The language module also has the functionality to generate multimodal utterance in respond to user command. When utterance request is received from the *Behaviours Generation* module, the language module performs the similar syntactical and semantics operations on the generated utterance to produce co-verbal gestures, which are then transmitted to the *Motion Generation* module for further processing. The current implementation of the system only deals with the iconic gesture for co-verbal behaviours.

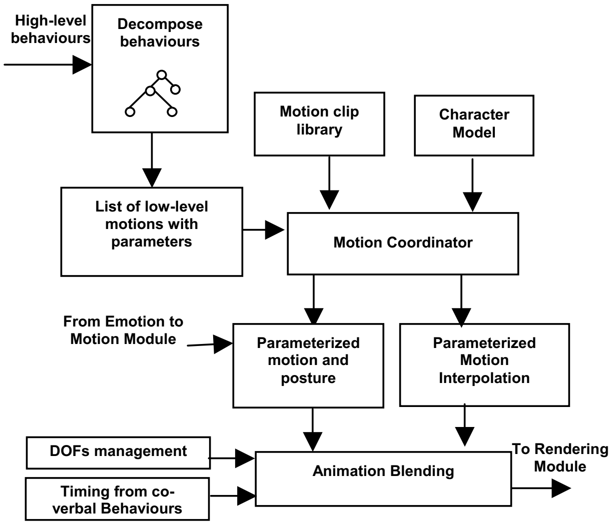

High-level behaviours
Decompose behaviours
Motion clip library
Character Model
List of low-level motions with parameters
Motion Coordinator
From Emotion to Motion Module
Parameterized motion and posture
Parameterized Motion Interpolation
DOFs management
Timing from co-verbal Behaviours
Animation Blending
To Rendering Module

Figure 2. Architecture and functionalities of the motion generation module

The *Motion Generation* module firstly perform the behaviours coordination and scheduling on the two suggested behaviours channels, namely selected body behaviours and co-verbal behaviours. For the co-verbal behaviours, timing is important as the speech accompanying gesture normally follows a preparation, stroke, and retraction pattern as suggested in [22]. Therefore, it is important to use the timing information in such gestures in order to seamlessly incorporated them into the overall body behaviours. The *Motion Generation* module then decomposes those behaviours into low-level motions with control parameters that can be realized by the motion clips library. The *Motion Coordinator*, upon receiving motions with parameters, retrieves base motions from the motion clips library and joints from the character skeleton that are required to perform the motions. Base motions are primarily produced by motion interpolation. On top of these base motions, parameterized motions are generated based on outputs from the *Emotion to Motion Module*. Such outputs are normally motion control parameters that are used to alter the way the base motions are animated or parameters to change the posture of the agent. All these generated motions are finally blended together to produce the final motions for the agent. The DOFs management assure that motion blending does not violate the limits of character's joints therefore avoiding un-natural motion. The functionalities of this *Motion Generation* module can be extended by incorporating new developed components. For instance, a physically-based modelling based motion module can be incorporated to increase the motion realism. In such cases, the adjustments (e.g. the update to involved DOFs) from this module can be blended into the existing motions via animation blending and DOF management.

Rendering module is responsible for displaying both of character animation and the virtual world onto the screen. It receives animation requests from the *Motion Generation module* and activates corresponding animation procedures with control parameters. The major challenge of designing such module is to achieve a balance between visual realism and the controllability of the animated 3D agent. The skinned mesh animation algorithm is used in the system [11].

3. Implementation and Results

The system was implemented in DirectX with MFC on a PC platform. The initial 3D human model contains around 6000 polygons and has around 23 degree of freedoms. The agent, for simplicity and experimental purpose, has 1 internal state (*Energy*) and two emotional states (*Happy, Angry*). The agent has base motions, namely *Walking, Running, and Resting*. It also has three parameterized motions, namely *HeadMovement(), PostureChange(), and ArmMovement()*. Its motions are arranged into a hierarchical actions network. Agent changes its base motions based on the fuzzy internal state *Energy*. Parameterized motions are generated based on the two emotional states, *Happy* and *Angry*.

Figure 3 shows that the agent is in searching state. In this state, the agent is given the task to capture an object controlled by the user. Its emotional states are updated based on dynamic events in the environment. In this case, the failure or success to capture the object has impact on his two emotional states. Emotion states are also updated according to time passing. This simulation used a default personality profile.

As shown in Figure 3, the events of failure or success to capture the object is evaluated and used to update the agent's emotional states, which subsequently resulting in different way of exhibiting its motions. The *Motion Generation module* produces parameterized motions and subsequently blends into the existing base motion animation sequences to achieve non-repetitive behaviours without the need to explicitly model such different style of animations in advance in modelling packages. The user is able to give commands to the agent and alter the agent's internal and emotional states through the user interface, observing the instant change of agent behaviours. Frame rate of the above simulation achieves an average of 100 based on a machine with *Pentium IV 2.8 GHz* CPU and a *Geforce 2 MX 400* graphics card.

4. Conclusion and Future Work

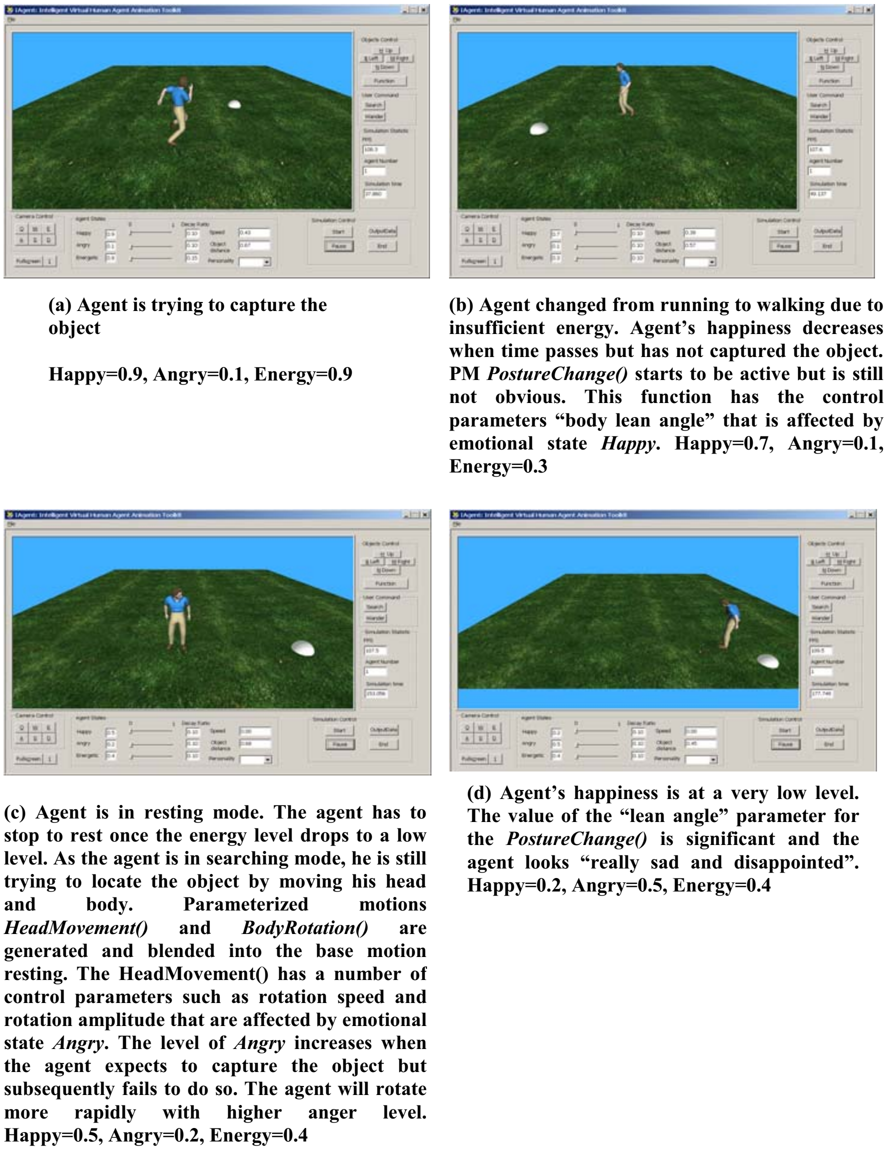
This paper has presented the design and implementation of a real time animation framework in which animated virtual human is capable of performing multimodal interactions with human user. The agent has the ability to capture information from its environment and determine what actions should be taken based on its *behaviours module*. The generated behaviours include body motion, posture, utterance and co-verbal gestures. An efficient *Motion Generation module* is developed to realize the selected behaviours and produce parameterized motion along with pre-generated animation sequence depending on agent's momentary emotional states. Furthermore, as the *Motion Generation module* has the control of the character model to a degree of freedom level and a flexible animation blending component, it is possible to integrate various existing motion planning, kinematics, and physically based motion into the framework in order to increase the realism of the human agent's motions. Future work can be enhanced by utilising more sophisticated behaviours processing paradigm to manage the coordination of various verbal and non-verbal behaviours in order to more effectively convey agent's communication goals and emotions. Furthermore, the incorporation of other types of co-verbal behaviours such as beats and metaphorical gesture into the system could further increase the believability of the human agent's behaviours.

5. References

[1] Burke, R., Isla, D., Downie, M., Ivanov,Y., Blumberg, B., "Creature Smarts: The Art and Architecture of a Virtual Brain", *In Proc. of the Game Developers Conference* (San Jose, CA, 2001), pp. 147-166.

[2] Noser, H. and Thalmann, D., "A Rule-based Interactive Behavioural Animation System for Humanoids", *IEEE Trans. On Visualization and Computer Graphics*, Vol. 5, No. 4, pp. 281-307.

[3] Terzopoulos, D., Tu, X., and Grzeszczuk, R., "Artificial Fishes: Autonomous Locomotion, Perception, Behavior, and Learning in a Simulated Physical World," *Journal of Artificial Life,* Vol. 1, No. 4., 1994, pp.327-351.



(a) Agent is trying to capture the object

Happy=0.9, Angry=0.1, Energy=0.9

(b) Agent changed from running to walking due to insufficient energy. Agent's happiness decreases when time passes but has not captured the object. PM *PostureChange()* starts to be active but is still not obvious. This function has the control parameters "body lean angle" that is affected by emotional state *Happy*. Happy=0.7, Angry=0.1, Energy=0.3

(c) Agent is in resting mode. The agent has to stop to rest once the energy level drops to a low level. As the agent is in searching mode, he is still trying to locate the object by moving his head and body. Parameterized motions *HeadMovement()* and *BodyRotation()* are generated and blended into the base motion resting. The HeadMovement() has a number of control parameters such as rotation speed and rotation amplitude that are affected by emotional state *Angry*. The level of *Angry* increases when the agent expects to capture the object but subsequently fails to do so. The agent will rotate more rapidly with higher anger level. Happy=0.5, Angry=0.2, Energy=0.4

(d) Agent's happiness is at a very low level. The value of the "lean angle" parameter for the *PostureChange()* is significant and the agent looks "really sad and disappointed". Happy=0.2, Angry=0.5, Energy=0.4

Figure 3. Agent in searching mode (emotional states influence action)

[4] Thalmann, D. and Monzani, J., "Behavioural Animation of Virtual Humans: What Kind of Law and Rules", *Proc. Computer Animation 2002*, IEEE CS Press, 2002, pp.154-163.

[5] Bruderlin, A., and Williams, A., "Motion signal processing", *Computer Graphics (Proc. of SIGGRAPH 95)*, pp. 97–104.

[6] Rose, C., Cohen, M.F., and Bodenheimer "Verbs and adverbs: Multidimensional Motion Interpolation", *IEEE Computer Graphics and applications*, September-October 1998 (Vol.18, No.5), pp. 32-40

[7] Kendon, A., "Gesticulation and Speech: Two Aspects of the Process of Utterance," in *The Relationship of Verbal and Nonverbal Communication*, M. R. Key, Ed., The hague: Mouton Publishers, pp. 207-227.

[8] Cassell, J., Vilhjalmsson, H. & Bickmore, T., "BEAT: the Behavior Expression Animation Toolkit". In *Proc. SIGGRAPH 2001*, pp. 477–486.

[9] Kopp, S. & Wachsmuth, I., "Synthesizing Multimodal Utterances for Conversational Agents", *Computer Animation and Virtual Worlds*: 15(1), 2004, pp. 39-52.

[10] Chi, D., Costa, M., Zhao, L., and Badler, N., “The EMOTE Model for Effort and Shape”, In *Proc. SIGGRAPH 2000, Computer Graphics Annual Conference* (New Orleans, Louisiana, 23-28 July, 2000), pp. 173-182.

[11] Wen, Z., Mehdi, Q, and Gough, N., “A New Animation Approach for Visualizing Intelligent Agent Behaviours in a Virtual Environment,” In *Proceedings of the Information Visualization* (London, UK, Jul.), pp.93-98.

[12] Bates, J., “The Role of Emotion in Believable Agents.” *Communications of the ACM 37*, no.7, pp.122-125.

[13] Blumberg, B.,“Action-selection in Hamsterdam: Lessons from Ethnology.” In *Proceedings of SAB'94* (Brighton, UK, Aug.8-12), pp.108-117.

[14] Oatley, K. and Johnson-Laird, P.N.,“Towards a Cognitive Theory of Emotions.” *Cognition and Emotion* 1, No.1, 1987, pp.29-50.

[15] Ortony, A., Clore, G.L. and Collins., *The Cognitive Structure of Emotions*. Cambridge University Press, 1988, UK.

[16] Bartneck, C., “Integrating the OCC Model of Emotions in Embodied Characters.” *In Proceedings of the on Virtual Conversational Characters: Applications, Methods, and Research Challenges* (Melbourne, Australia, 2002).

[17] Ball, G. and Breese, J., “Emotion and Personality in a Conversational Character”. In *Proceedings of the Workshop on Embodied Conversational Character* (California, USA, Oct, 1998), pp.83-84.

[18] Marsella, S. and Gratch, J., “A Step Toward Irrationality: Using Emotion to Change Belief.” In *Proceeding of First International Joint Conference on Autonomous Agents and Multi-Agent Systems* (Bologna, Italy, Jul. 15-19,2002), pp.334-341.

[19] Costa, P. T. and McCrae, R. R., “Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory.” *Psychological Assessment* 4, 1992, pp.5–13.

[20] Egges, A., Kshirsagar, S. and Magnenat-Thalmann, N., "Generic Personality and Emotion Simulation for Conversational Agents.” *Computer Animation and Virtual Worlds* 15, no.1 (Jan.): 1-13. *Proceedings of 11th International Conference in Intelligent Systems* (Arlington , USA, Jun. 13-15, 2004), pp.47-50.

[21] Dennis Grinberg, John Lafferty and Daniel Sleator., “A Robust Parsing Algorithm for Link Grammars”. Carnegie Mellon University Computer Science technical report CMU-CS-95-125, and *Proceedings of the Fourth International Workshop on Parsing Technologies*, Prague, September, 1995.

[22] McNeil, D., *Hand and Mind: What Gestures Reveal about Thought*, 1992, University of Chicago Press.