Nonverbal Behavior Generator for Embodied Conversational Agents

Jina Lee and Stacy Marsella

University of Southern California
Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina Del Rey, CA 90292
{jinal, marsella}@isi.edu

Abstract. Believable nonverbal behaviors for embodied conversational agents (ECA) can create a more immersive experience for users and improve the effectiveness of communication. This paper describes a nonverbal behavior generator that analyzes the syntactic and semantic structure of the surface text as well as the affective state of the ECA and annotates the surface text with appropriate nonverbal behaviors. A number of video clips of people conversing were analyzed to extract the nonverbal behavior generation rules. The system works in real-time and is user-extensible so that users can easily modify or extend the current behavior generation rules.

1 Introduction

Nonverbal behaviors serve to repeat, contradict, substitute, complement, accent, or regulate spoken communication [1]. They can include facial expressions, head movements, body gesture, body posture, or eye gaze. Nonverbal behaviors can also be affected by a range of affective phenomena. For example, an angry person might display lowered eyebrows and tensed lips and more expressive body gestures than one who is not. Such behavior can in turn influence the beliefs, emotions, and behavior of observers.

Embodied conversational agents (ECA) with appropriate nonverbal behaviors can support interaction with users that ideally mirrors face-to-face human interaction. Nonverbal behaviors also can help create a stronger relationship between the ECA and user as well as allow applications to have richer, more expressive characters. Overall, appropriate nonverbal behaviors should provide users with a more immersive experience while interacting with ECAs, whether they are characters in video games, intelligent tutoring systems, or customer service applications [2].

This paper describes our approach for creating a nonverbal behavior generator module for ECAs that assigns behaviors to the ECA's utterances. We are especially interested in an approach that generates nonverbal behaviors provided only the surface text and, when available, the ECA's emotional state, turn-taking strategy, coping strategy, and overall communicative intent. In general, we seek

J. Gratch et al. (Eds.): IVA 2006, LNAI 4133, pp. 243–255, 2006.

Fig. 1. SASO's SmartBody

a robust process that does not make any strong assumptions about markup of communicative intent in the surface text. Often such markup is not available unless entered manually. Even in systems that use natural language generation to create the surface text (e.g., Stabilization and Support Operations system [3]), the natural language generation may not pass down detailed information about how parts of the surface text (a phrase or word, for example) convey specific aspects of the communicative intent or emotional state. As a result, the nonverbal behavior generator often lacks sufficiently detailed information and must rely to varying degrees on analyzing the surface text. Therefore, a key interest here is whether we can extract information from the lexical, syntactic, and semantic structure of the surface text that can support the generation of believable nonverbal behaviors.

Our nonverbal behavior generator has been incorporated into SmartBody, an ECA developed at University of Southern California [1]. SmartBody project is part of the Stabilization and Support Operations (SASO) research prototype, which grew out of the Mission Rehearsal Environment [3] to teach leadership and negotiation skills under high stress situations. In this system, the trainees interact and negotiate with life-size ECA that reside in a virtual environment. Figure 1 shows SmartBody, in this case a doctor, whom the trainee interacts with.

The next section describes related works. Section three describes research on nonverbal behavior and our analysis of video clips to derive the nonverbal behavior generation rules. Section four describes the system architecture of the nonverbal behavior generator and an example that walks through the behavior generation process. We also discuss the extensibility of the nonverbal behavior generator and propose directions for future work.

[1] This is a joint work of USC Information Sciences Institute and USC Institute for Creative Technologies.

2 Related Work

Mirroring the studies of nonverbal behavior in human communication, ECA research has shown that there is a significant improvement in the user's level of engagement while interacting with ECA that displayed believable nonverbal behaviors. The work of Fabri et al. [2] suggests that ECA with expressive abilities can increase the sense of togetherness or community feeling. Durlach and Slater [4] observed that ECA with even primitive nonverbal behaviors generate strong emotional responses from the users.

The effort to construct expressive ECA ranges from animating human faces with various facial expressions to generating complex body gestures that convey emotions and communicative intent. Rea [5] engages in a face-to-face interaction with a user and models the intention and communicative intention of the agent to generate appropriate facial expressions and body gestures. Becheiraz and Thalmann [6] developed a behavioral animation system for virtual characters by modeling the emergence conditions for each character's personality and intentions. Striegnitz et al. [7] developed an ECA that autonomously generates hand gestures while giving directions to the user.

There has also been work that emphasizes the reusability of the nonverbal behavior generators by separating the concept of behavior generation and behavior realization. The BEAT [8] system is a plug-in model for nonverbal behavior generation that extracts the linguistic structure of the text and suggests appropriate nonverbal behaviors. It allows users to add new entries to extend the gesture library or modify strategies for generating or filtering out the behaviors.

BEAT's functions and purpose very much informed our work; however, there are several differences. We are crafting our system around the new BML and FML standards [9]. This should provide a clearer, more general and standardized interface for communicative intent and behavior specification. BEAT had a variety of pre-knowledge about the surface text to be delivered at different abstraction levels, which is not the case in our nonverbal behavior generator. We are interested in exploring the degree to which nonverbal behavior generator can work only with the surface text and a minimal set of specification on the communicative intent at a high level of abstraction such as the turn-taking information and the affective state. We are also exploring a different range of expressive phenomena that is complementary to BEAT's work. Specifically, we are analyzing videos of emotional dialogues. Finally, BEAT included a commercial language tagger, while we are planning to maintain our nonverbal behavior generator open-source.

3 Study of Nonverbal Behaviors

3.1 Nonverbal Behaviors and Their Functionalities

There is a large research literature on the functionalities of nonverbal behaviors during face-to-face communication [10] [11] [12] [13] [14]. Heylen [12] summarizes the functions of head movements during conversations. Some included are:

to signal yes or no, enhance communicative attention, anticipate an attempt to capture the floor, signal the intention to continue, mark the contrast with the immediately preceding utterances, and mark uncertain statements and lexical repairs. Kendon [13] describes the different contexts in which the head shake may be used. Head shake is used with or without verbal utterances as a component of negative expression, when a speaker makes a superlative or intensified expression as in 'very very old', when a speaker self-corrects himself, or to express doubt about what he is saying. In [14], lateral sweep or head shakes co-occurs with concepts of inclusivity such as 'everyone' and 'everything' and intensification with lexical choices such as 'very', 'a lot', 'great', 'really'. Side-to-side shakes also correlate with expressions of uncertainty and lexical repairs. During narration, head nods function as signs of affirmation and backchannel requests to the speakers. Speakers also predictably change the head position for alternatives or items in a list. Ekman [10] describes eyebrow movements for emotional expressions and conversational signals. Some examples are eyebrow raise or frowning to accent a particular word or to emphasize a particular conversation point. One of the goals for our nonverbal behavior generator is to find features in the dialogue that convey these attributes and annotate them with appropriate nonverbal behaviors that are consistent with the research literature. Although the above discussion is couched in general terms, nonverbal behaviors vary across cultures and even individuals. We return to this issue later.

3.2 Video Data Analysis

In addition to the existing research literature, we have also studied the uses of nonverbal behaviors in video clips of people conversing. The literature is useful for broadly classifying the behaviors. However, to better assess whether it is feasible to build behavior generation rules that could map from text to behavior, an analysis of actual conversations was needed.

We obtained video clips of users interacting with the Sensitive Artificial Listener system from the Human-Machine Interaction Network on Emotion [15]. Sensitive Artificial Listener (SAL) is a technique to engage users in emotionally colored interactive discourse [16]. SAL is modeled on an ELIZA scenario [17], a computer emulation of a psychotherapist. In SAL, the operator plays the role of one of four characters with different personalities and responds to the user with pre-defined scripts. The main goal is to pull the user's emotion towards the character's emotional state.

17 video clips were analyzed, each ranging from five to ten minutes in length. The video clips capture only the users' torso and above, and we mainly annotated the facial expressions and head movements exhibited by the users. For each video clip, we annotated the types of nonverbal behaviors portrayed, their frequency, time frame, spoken utterance, and the users' emotional states when the behavior occurred. This was documented in an XML form for easy parsing and processing.

There were a number of different nonverbal behaviors observed in these video clips. These behaviors include:

– Head Movement: nods, shakes, head moved to the side, head tilt, pulled back, pulled down
– Eyebrow Movement: brow raised, brow lowered, brow flashes
– Eye/Gaze Movement: look up, look down, look away, eyes squinted, eyes squeezed, eyes rolled
– Others: shoulder shrug, mouth pulled on one side

To annotate the utterances, we adopted the labels used in the literature and created a few more for the utterances in which we observed a nonverbal behavior but no appropriate labels were used in the literature. The labels used are affirmation, negation, contrast, intensification, inclusivity, obligation, listing, assumption, possibility, response request, and word search. For each utterance accompanying nonverbal behaviors, we attached the labels applicable to the utterance and annotated the behaviors. There were 161 utterances that were annotated using these labels. Table 1 shows the distribution of the number of utterances that includes each label.

Table 1. Breakdown of the number of utterances with corrsponding labels

Label	# of utterances (out of 161)	Label	# of utterances (out of 161)
Affirmation	39	Response Reqeust	9
Negation	62	Inclusivity	7
Intensification	41	Obligation	6
Word Search	25	Assumption	3
Contrast	9		

A number of utterances were annotated with two or more labels, which is why the sum of each component exceeds 161. Besides these 161 utterances, there were 58 utterances that accompanied nonverbal behaviors but could not be labeled appropriately because there was not a clear and consistent pattern between the utterance and the behaviors. The nonverbal behaviors on these utterances were usually observed at the beginning of the sentence or when the user was emphasizing a particular word or context, but the behaviors varied in each case.

In general, we found a close match between the literature and our video analysis on the mappings of nonverbal behaviors to certain utterances. For example, a head shake usually occurred when a word with inclusive meaning such as 'all' and 'everything' was spoken and lowered eyebrow with a head nod or shake occurred when intensifying words like 'really' was spoken. We also analyzed the parse trees of the utterances and found mappings between certain behaviors and syntactic structures. Interjections, which were usually associated with the words 'yes', 'no', and 'well' in the video clips accompanied either a head nod, shake, or tilt in most cases.

Based on the study from the literature and our video analysis, we created a list of nonverbal behavior generation rules, which are described in Figure 2.

(1) INTERJECTION: Head nod, shake, or tilt co–occurring with these words:
- *Yes, no, well*

(1) NEGATION: Head shakes and brow frown throughout the whole sentence or phrase these words occur:
- *No, not, nothing, can't, cannot*

(2) AFFIRMATION: Head nods and brow raise throughout the whole sentence or phrase these words occur:
- *Yes, yeah, I do, I am, We have, We do, You have, true, OK*

(3) ASSUMPTION / POSSIBILITY: Head nods throughout the sentence or phrase and brow frown when these words occur:
- *I guess, I suppose, I think, maybe, perhaps, could, probably*

(3) OBLIGATION: Head nod once co–occurring with these words:
- *Have to, need to, ought to*

(4) CONTRAST: Head moved to the side (lateral movement) and brow raise co–occurring with these words:
- *But, however*

(4) INCLUSIVITY: Lateral head sweep co–occurring on these words:
- *Everything, all, whole, several, plenty, full*

(4) INTENSIFICATION: Head nod and brow frown co–occurring with these words:
- *Really, very, quite, completely, wonderful, great, absolutely, gorgeous, huge, fantastic, so, amazing, just, quite, important, . . .*

(4) LISTING: Head moved to the side (lateral movement) and to the other before and after the word 'and':
- *X and Y*

(4) RESPONSE REQUEST: Head moved to the side and brow raise co-occurring with these words:
- *You know*

(4) WORD SEARCH: Head tilt, brow raise, gaze away co-occurring with these words:
- *Um, uh, well*

Fig. 2. Nonverbal behavior generation rules. The numbers in the parenthesis indicates the priority or each rule.

Each rule has associated nonverbal behaviors and a set of words that are usually spoken when the nonverbal behavior is exhibited. We also defined a priority value for each rule based on our analysis to resolve conflicts between rules that could co-occur. For example, in the utterance 'Maybe we shouldn't do that', both the assumption rule and the negation rule could be applied. However, the video analysis tells us that the negation rule overrides the assumption rule in those cases. In general, the nonverbal behavior rules that occur over the whole sentence or phrases overrule those that occur on a single word.

Following are examples on how the rules are applied to given surface texts.

Example 1

Surface Text:

I do, I do. I'm looking forward to that but I can't rest until I get this work done.

Rules applied:

Affirmation rule from *I do* and *I'm*

Negation rule from *can't*

(Contrast rule applied from *but* is overridden by the negation rule)

Nonverbal Behaviors:

Head nods on *I do, I do* and *I'm looking forward*

Head shakes on *I can't rest*

Example 2

Surface Text:

Yes, Prudence, many times. I actually quite like you.

Rules applied:

Interjection rule from *yes*

Intensification rule from *quite*

Nonverbal Behaviors:

Head nod on *yes*

Head nod on *quite*

In addition to the nonverbal behaviors associated with certain dialogue elements, we also put small head nods on phrasal boundaries. This is based on our experience that it makes the ECA more life-like, perhaps because the human head is often in constant (small) motion as a person talks.

The next section describes how we use these rules to create execution commands for believable nonverbal behaviors.

4 System Architecture

4.1 Overview

The nonverbal behavior generator is built to be modular and to operate in real time with user-extensible behavior generation rules. The input and output interaction to the system is done by a message pipeline system, and the main data structure for the inputs and outputs is in XML form. More specifically, we are using Function Markup Language (FML) and Behavior Markup Language (BML) as part of the input and output messages (see the next section for more details on FML and BML). The nonverbal behavior generator uses two major tools to select and schedule behaviors: a natural language parser and an XML

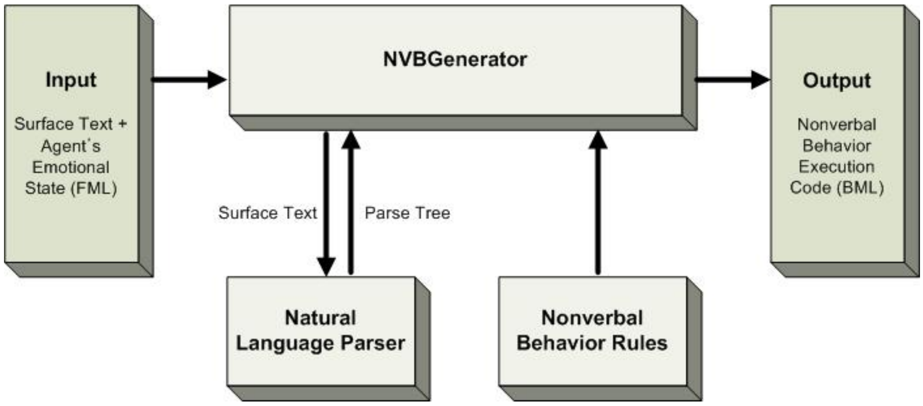

Fig. 3. System architecture of the nonverbal behavior generator

stylesheet transformation (XSLT) processor. XSL is a language to transform XML documents into other XHTML documents or XML documents. In our case, we will be transforming the input XML string by inserting time markers to the surface utterance and behavior execution codes. The nonverbal behavior generation rules are also represented in XSL format.

Figure 3 illustrates the overview of the system's structure. The nonverbal behavior generator's input XML string contains the surface text of the agent as well as other affective information such as the agent's emotional state, emphasis point, and coping strategy. The NVBGenerator module parses this XML message, registers the agent's affective information, and extracts the surface text. The surface text is then sent to the natural language parser to obtain the syntactic structure of the utterance. Given the parsed result of the utterance and the behavior generation rules, the NVBGenerator selects the appropriate behavior(s). The selected behaviors are then customized and modified by the affective information of the agent. Finally, the execution code for the chosen behavior(s) are generated and sent out to the virtual human controller. The following sections describe parts of the processing steps in greater detail.

4.2 Function Markup Language and Behavior Markup Language

The Social Performance Framework [9] [18] and more recently SAIBA [19] are being developed to modularize the design and research of embodied conversational agents. These frameworks define modules that make clear distinction between the communicative intent and behavior descriptions of the ECA with XML based interfaces. This distinction is defined by two markup languages FML and BML, which consolidate a range of prior work in markup languages (such as the Affective Presentation Markup Language [20] and Multimodal Utterance Representation Markup Language [21]). Function Markup Language (FML) specifies the communicative and expressive intent of the agent and will be part of the input message to our nonverbal behavior generator. The following describes some of the elements defined in FML.

- AFFECT: The affective state of the speaker (e.g. JOY, DISTRESS, RESENTMENT, FEAR, ANGER...).
- COPING: Identification of a coping strategy employed by the speaker.
- EMPHASIS: Speaker wants listeners to pay particular attention to this part of the spoken text.
- TURN: Management of speaking turns (TAKE, GIVE, KEEP).

Behavior Markup Language (BML), on the other hand, describes the verbal and nonverbal behaviors an agent will execute. The elements of BML roughly correspond to the parts of human body and the attributes of each element further define the details of specific behavior execution information such as the start and end time and the frequency of the behavior. The set of elements defined in BML includes,

- HEAD: Movement of the head independent of eyes.
- FACE: Movement in the face.
- GAZE: Coordinated movement of the eyes, neck, and torso, indicating where the character is looking.
- BODY: General movement of the body.
- GESTURE: Coordinated movement with arms and hands.
- SPEECH: Spoken delivery.
- LIPS: Movement of the mouth.
- ANIMATION: Plays back a character animation clip.

The selected behaviors from our nonverbal behavior generator are encoded using these BML tags and be included in the output message. Incorporating FML and BML to specify the communicative intent and the nonverbal behaviors of the agent not only gives the structural format to express these information, but allows the developer to easily process the information using any XML processor, which is widely available.

4.3 Nonverbal Behavior Generation Process

Let's have a closer look at how the nonverbal behaviors are selected and generated. Assume the input message to the generator contains the following information.

- Surface text:
  Yes, I completely agree. I am not interested only in myself, you know.
- Emphasis: Emphasis on *myself*
- Affect: Neutral

The NVBGenerator first parses the input message, extracts the surface string, and sends it to the natural language parser. We are currently using Charniak's

Fig. 4. Nonverbal behaviors animated on SmartBody

parser [22] to process the utterance. The parse tree is sent back and the NVB-Generator inserts time markers between every word of the utterance. Then the NVBGenerator analyzes the semantic and syntactic structure of the utterance to decide which rules could be fired and inserts XML tags for such rules. The XSLT processor looks at these rule tags and matches them to insert the BML codes into the output message. But if there are two rules that overlap with each other, the one with a higher priority will be selected.

In the example above, the rules that apply to the given surface text will be, *interjection rule*, which creates BML codes for a head nod on the word 'Yes', *intensification rule*, which puts a head nod and lowered brow movement on the word 'completely', *negation rule*, which puts head shakes on 'I am not interested', *first noun phrase rule*, which puts a small head nod after 'myself', and the *response request rule*, which puts head nod after 'you know'. Since there is an emphasis on the word 'myself', the NVBGenerator will replace the medium head nod to a big nod and insert lowered brow movement when 'myself' is spoken. The SmartBody system also has a number of pre-animated gesture clips that could be used in place of the BML codes. For example, we have an animation clip where the ECA puts his hand up and shakes his head, which could be used when the *negation rule* is selected instead of outputting a BML code for head shake. Figure 4 shows examples of some nonverbal behaviors animated on SmartBody. Finally, the output message consisting of the surface text with time markers and BML codes are sent to the SmartBody controller [23] that synchronizes and animates the nonverbal behaviors.

4.4 Extensibility and Specialization

The nonverbal behavior generator has been designed for easy extension for the users. As mentioned in section 4.1, the nonverbal behavior generation rules are represented in XSL format. There is one file that stores the behavior descriptions for different nonverbal behaviors and another file that stores the association between the rules and the nonverbal behaviors. More specifically, the behavior description file stores the BML codes for different behaviors such as *big_head_nod,* *small_head_shake*, and *brow_frown* and the behavior generation rule file stores the information on which behaviors should be generated for each rule. For example, when intensification rule is applied, a small head nod and brow frowning should occur. As described in section 4.2, the whole behavior generation process is done in three steps; first the NVBGenerator analyzes the surface text and inserts an XML tag for the appropriate rule. Then the behavior generation rule file matches this tag to see which behaviors should occur, and finally the appropriate BML codes stored in the behavior description file is inserted to the output message.

The separation between behavior descriptions and nonverbal behavior generation rules allows easy modification and extension without affecting one another. For example, it is simple to add new entries of gesture animations or behavior descriptions into the system. As the animator creates new gesture animations or a programmer creates a new procedural behavior, one can simply extend the behavior description file to add the name of the animations or behavior description for future use. It is also easy to modify the rules that invoke the behavior descriptions. For example, if the current rule for *inclusivity* contains a lateral head movement but one wishes to add a brow raise to it, he or she simply needs to add lines to the file storing the behavior generation rules, which will call the behavior description for brow raise. This separation also supports supports specialization of behavior according to individual or cultural traits. For example, we can have different rules for inclusivity based on culturally-specific gesturing tendencies.

Using XSL to represent the behavior descriptions and behavior generation rules also allows the user to make modifications without knowing the details of the nonverbal behavior generator. There is no need to have other programming language skills or study how the behavior generator is implemented. By learning simple patterns on how to add XSLT templates, one can create, modify or delete behavior descriptions and rules.

5 Conclusion and Future Work

We have developed a framework for text-to-speech nonverbal behavior generation. It analyzes the syntactic and semantic structure of the input text and generates appropriate head movements, facial expressions, and body gestures. We studied a number of video clips to develop rules that map specific words, phrases, or speech acts and constructed our behavior generation rules according to this. The behavior generator is designed to be easy for users to modify or create behavior descriptions and behavior generation rules. The module was successfully incorporated into the SASO and SmartBody system, using the SAIBA markup structure, and works in real time. It has also been fielded in a cultural training application being developed at the Institute for Creative Technology.

Much work still remains to improve the system. Our next step would be to evaluate the system and the behaviors generated. We are particularly interested in the user's responses to the behaviors and what they infer from the behaviors. We expect our current rules are too limited and overly general in their applicability. Thus, we are also seeking ways to use various machine learning techniques to aid us in the process of rule generation. One straightforward approach would be to learn the mapping between bigrams or trigrams of words to gestures. This would require a large gesture corpora; however a suitable corpora for our work is currently not available. In the absence of a large corpora, we rather expect the learning should be informed by higher level features such as syntactic, lexical, and semantic structure of the utterance or the ECA's emotional state, similar to what we used to craft the rules by hand.

Furthermore, we would like to modify the nonverbal behavior generation given the information on ECA's supposed gender, age, culture, or personality. The system also lacks a good knowledge base of the environment in which the ECA resides. A tight connection to the knowledge base of the objects and agents in the virtual world will allow the ECA to have more sophisticated behaviors such as deictic gestures that correctly points at the object referred. Finally, we would like to model the affective state of the user interacting with the ECA and generate appropriate behaviors that respond not only to agent's emotions but also to the user's emotions.

Acknowledgments

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. Knapp, M., Hall, J.: Nonverbal Communication in Human Interaction. 4th edn. Harcourt Brace College Publishers (1997)
2. Fabri, M., Moore, D., Hobbs, D.: Expressive agents: Non-verbal communication in collaborative virtual environments. In: Proceedings of Autonomous Agents and Multi-Agent Systems, Bologna, Italy (2002)
3. Swartout, W., Hill, R., Gratch, J., Johnson, W., Kyriakakis, C., Labore, K., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiebaux, M., Tuch, L., Whitney, R.: Toward the holodeck: Integrating graphics, sound, character and story. In: Proceedings of 5th International Conference on Autonomous Agents, Montreal, Canada (2001)
4. Durlach, N., Slater, M.: Presence in shared virtual environments and virtual togetherness. In: BT Workshop on Presence in Shared Virtual Environments, Ipswich, UK (1998)
5. Cassell, J., Vilhjálmsson, H., Chang, K., Bickmore, T., Campbell, L., Yan, H.: Requirements for an architecture for embodied conversational characters. In Magnenat-Thalmann, N., Thalmann, D., eds.: Computer Animation and Simulation '99. Springer Verlag, Vinna, Austria (1999) 109–120
6. Becheiraz, P., Thalmann, D.: A behavioral animation system for autonomous actors personified by emotions. In: Proceedings of the 1st Workshop on Embodied Conversational Characters (WECC), Lake Tahoe, CA (1998) 57–65

7. Striegnitz, K., Tepper, P., Lovett, A., Cassell, J.: Knowledge representation for generating locating gestures in route directions. In: Proceedings of Workshop on Spatial Language and Dialogue, Delmenhorst, Germany (2005)
8. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: The behavior expression animation toolkit. In: Proceedings of ACM SIGGRAPH, New York, NY, ACM Press / ACM SIGGRAPH (2001) 477–486
9. Vilhjálmsson, H., Marsella, S.: Social performance framework. In: Workshop on Modular Construction of Human-Like Intelligence at the AAAI 20th National Conference on Artificual Intelligence, Pittsburgh, PA (2005)
10. Ekman, P.: About brows: emotional and conversational signals. In von Cranach, M., Foppa, K., Lepenies, W., Ploog, D., eds.: Human Ethology. Cambridge University Press (1979) 169–248
11. Hadar, U., Steiner, T., Grant, E., Clifford Rose, F.: Kinematics of head movement accompanying speech during conversation. Human Movement Science **2** (1983) 35–46
12. Heylen, D.: Challenges ahead. In: Proceedings of AISB Symposium on Social Virtual Agents. (in press)
13. Kendon, A.: Some uses of head shake. Gesture (2) (2003) 147–182
14. McClave, E.: Linguistic functions of head movements in the context of speech. Journal of Pragmatics (32) (2000) 855–878
15. The HUMAINE Consortium: The HUMAINE portal. Retrieved April 7, 2006, from http://emotion-research.net/ (2006)
16. The HUMAINE Consortium: Multimodal data in action and interaction: a library of recordings and labelling schemes. Retrieved April 14, 2006, from http://emotion-research.net/deliverables/ (2004)
17. Weizenbaum, J.: ELIZA – a computer program for the study of natural language communication between man and machines. Communications of the Association for Computing Machinery **9** (1996) 36–45
18. n.a.: Behavior markup language (BML) specification. Retrieved June 6, 2006, from http://twiki.isi.edu/Public/BMLSpecification/ (2006)
19. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thorisson, K., Vilhjálmsson, H.: Towards a common framework for multimodal generation in embodied conversation agents: a behavior markup language. In: International Conference on Virtual Agents, 2006, Marina del Rey, CA (submitted)
20. DeCarolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a mark-up language for believable behavior generation. In Prendinger, H., Ishizuka, M., eds.: Life-like Characters. Tools, Affective Functions and Applications. Springer (2004) 65–85
21. Kopp S., W.I.: Synthesizing multimodal utterances for conversational agents. Computer Animation and Virtual Worlds **15**(1) (2004) 39–52
22. Chariank, E.: A maximum-entropy-inspired parser. In: Proceedings of North American Chapter of the Association for Computational Linguistics. (2000)
23. Kallmann, M., Marsella, S.: Hierarchical motion controllers for real-time autonomous virtual humans. In: Proceedings of the 5th International working conference on Intelligent Virtual Agents (IVA), Kos, Greece (2005) 243–265