Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Cerebella: Automatic Generation of Nonverbal Behavior for Virtual Humans

Margot Lhommet and **Yuyu Xu** and **Stacy Marsella**
Northeastern University
360 Huntington Avenue
Boston, MA, 02115

Abstract

Our method automatically generates realistic nonverbal performances for virtual characters to accompany spoken utterances. It analyses the acoustic, syntactic, semantic and rhetorical properties of the utterance text and audio signal to generate nonverbal behavior such as such as head movements, eye saccades, and novel gesture animations based on co-articulation.

Overview

The flip of a hand, a raising of an eyebrow, a gaze shift: a range of physical, nonverbal behaviors accompany speech in face-to-face interactions. They are pervasive in every moment of dialog and convey meaning that powerfully influence interaction.

Our interest in such behaviors lies in a desire to automate the generation of nonverbal behavior for convincing, life-like virtual character performances. Specifically, we demonstrate the automatic generation of a character's nonverbal behavior from the audio and text of the dialog they must speak.

Nonverbals can stand in different relations to the verbal content, serving a variety of functions in face-to-face interaction. Shifts in topic can be cued by shifts in posture or in head pose. Comparison and contrasts between abstract ideas can be emphasized by deictic gestures that point at the opposing ideas as if they each had a distinct physical location (McNeill 1992). The form of nonverbal gestures is often tied to physical metaphors; rejecting an idea can be illustrated by a sideways flip of the hand (Calbris 2011). A wide range of mental states and character traits can also be conveyed: gaze reveals thought processes, blushing suggests shyness and facial expressions convey emotions. Finally, nonverbals help manage the conversation, for example by signaling the desire to hold onto, get or hand over the dialog turn (Bavelas 1994).

Generating nonverbal behaviors must additionally take into account that they are synchronized with the dialog. For instance, the stroke of a hand gesture, a nod or eyebrow raise are often used to emphasize a word or phrase in the speech. Alteration of the timing will change what words are emphasized and consequently how the utterance is understood.

Such challenges make the pattern and timing of the behavior animations that accompany utterances unique to the utterance and the state of the character. Manual creation of the behaviors by hand animation and/or motion capture are consequently time consuming, costly, and require considerable expertise from the animator or the motion capture performer.

This has led us to research and develop Cerebella, an automatic system to generate expressive, life-like nonverbal behaviors (including nonverbal behaviors accompanying the speech, responses to perceptual events and listening behaviors). Cerebella is designed to operate in multiple modes. If the virtual character has mental processes that provide communicative functions, Cerebella will generate appropriate behaviors. However, in the absence of such information, Cerebella infers underlying communicative functions from the audio and text, and generates a performance by selecting appropriate animations and procedural behaviors.

This latter approach has illustrated its effectiveness in a variety of applications: the use as an embodied conversational agent (DeVault et al. 2014) or inside a previsualization tool for film (Marsella et al. 2013).

The interactive demonstration associated to this paper allows you to record your own lines and watch the virtual human perform them.

Generation Pipeline

A brief overview of the analyses is given bellow (for more detail, see (Marsella et al. 2013)):

1. Acoustic Processing: the spoken utterance is analyzed to derive information on what words are being stressed.
2. Syntactic Analysis: the sentence text is parsed to derive its syntactic structure.
3. Function Derivation: the utterance's communicative functions are inferred using forward-chaining rules to build up a hierarchically structured lexical, semantic, metaphoric and pragmatic analysis.
4. Behavior Mapping: a set of nonverbal behavior rules maps from communicative functions to classes of nonverbal behaviors.
5. Animation Specification: a schedule of behaviors is generated by mapping behavior classes to specific behaviors.

Mappings can be customized to support individual differences including personality, culture, gender and body types.

6. Animation Synthesis: the animation engine processes the schedule of behaviors and synthesizes the performance.

A central contribution of this work is the deep and novel types of analysis incorporated in a comprehensive, automated approach that can generate a full range of nonverbals. To cite a few, the system detects metaphors in the language to drive selection of metaphoric gestures (Lhommet and Marsella 2014). For example, events can have locations in space and abstract ideas can be considered as concrete objects, allowing them to have physical properties (like "a big idea") that can be reflected in gesture. Rhetorical structures like comparisons and contrasts suggest abstract deictic gestures (e.g., this idea as opposed to that idea can be conveyed gesturing left than right).

Human gesturing has a hierarchical structure that serves demarcative, expressive and referential purposes. Within a gesture performance, some features such as hand shape, movement or location in space, may be coupled across gestures while others serve at times a key role in distinguishing individual gestures and the meaning they convey. As opposed to approaches that focus more on realizing individual gesture, Cerebella considers the relation between gestures as part of an overall gesture performance by putting constraints on the gesture selection and realization algorithms (Xu, Pelachaud, and Marsella 2014).

Cerebella uses Smartbody (Thiebaux et al. 2008) a virtual character animation system that can address key requirements for realistic nonverbal behavior generation: gesture holds can provide emphasis, co-articulation can be realized between gestures to indicate a continuous ideational segment, and the hands may be forced into a rest position between gestures to indicate the end of an idea.

Demonstration

The demonstration illustrates how Cerebella automatically generates nonverbal behaviors for virtual humans. Users are invited to record a line of dialog of their choice. This audio file is processed by Cerebella to generate a nonverbal performance that illustrates the spoken line. Finally, the performance is realized by the virtual humans of the Virtual Human Toolkit (Hartholt et al. 2013)(see Figure 1).

Figure 1: Our demo uses the virtual humans from the Virtual Human Toolkit.

Acknowledgments

We are grateful to Teresa Dey for the art assets used in this work. This work is supported by the Air Force Office of Scientific Research (FA9550-09-1-0507) and the U.S. Army RDECOM. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

Bavelas, J. B. 1994. Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction* 27(3):201–221.

Calbris, G. 2011. *Elements of Meaning in Gesture*, volume 5. Amsterdam: John Benjamins Publishing.

DeVault; Artstein; Benn; Dey; Fast; Gainer; Georgila; Gratch; Hartholt; Lhommet; Lucas; Marsella; Morbini; Nazarian; Scherer; Stratou; Suri; Traum; Wood; Xu; Rizzo; and Morency. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems.

Hartholt, A.; Traum, D.; Marsella, S.; Shapiro, A.; Stratou, G.; Leuski, A.; Morency, L.-P.; and Gratch, J. 2013. All together now. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents*, 368–381. Springer.

Lhommet, M., and Marsella, S. 2014. Metaphoric gestures: Towards grounded mental spaces. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents*, 264–274. Springer.

Marsella, S.; Xu, Y.; Lhommet, M.; Feng, A.; Scherer, S.; and Shapiro, A. 2013. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 25–35. New York, NY, USA: ACM.

McNeill, D. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Thiebaux, M.; Marsella, S.; Marshall, A. N.; and Kallmann, M. 2008. SmartBody: behavior realization for embodied conversational agents. In *Proceedings of the 7th International Conference on Autonomous Agents and Multi-agent Systems*, 151–158. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Xu, Y.; Pelachaud, C.; and Marsella, S. 2014. Compound gesture generation: A model based on ideational units. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents*, 477–491. Springer.