Expressive Gesture Model for Humanoid Robot

Le Quoc Anh and Catherine Pelachaud

CNRS, LTCI Telecom ParisTech, France
{quoc-anh.le,catherine.pelachaud}@telecom-paristech.fr

Abstract. This paper presents an expressive gesture model that generates communicative gestures accompanying speech for the humanoid robot Nao. The research work focuses mainly on the expressivity of robot gestures being coordinated with speech. To reach this objective, we have extended and developed our existing virtual agent platform GRETA to be adapted to the robot. Gestural prototypes are described symbolically and stored in a gestural database, called lexicon. Given a set of intentions and emotional states to communicate the system selects from the robot lexicon corresponding gestures. After that the selected gestures are planned to synchronize speech and then instantiated in robot joint values while taking into account parameters of gestural expressivity such as temporal extension, spatial extension, fluidity, power and repetitivity. In this paper, we will provide a detailed overview of our proposed model.

Keywords: expressive gesture, lexicon, BML, GRETA, NAO.

1 Objectives

Many studies have shown the importance of the expressive gestures in communicating messages as well as in expressing emotions in conversation. They are necessary for speaker to formulate his thoughts [12]. They are also crucial for listeners. For instance they convey complementary, supplementary or even contradictory information to the one indicated by speech [9].

The objective of this thesis is to equip a physical humanoid robot with capability of producing expressive gestures while talking. This research is conducted within the frame of a project of the French Nation Agency for Research, ANR GVLEX that has started since 2009 and lasts for 3 years. The project aims to model a humanoid robot, NAO, developed by Aldebaran [4], able to read a story in an expressive manner to children for several minutes without boring them. In this project, a proposed system takes as input a text to be said by the agent. The text has been enriched with information on the manner the text ought to be said (i.e. with which communicative acts it should be said). The behavioral engine selects the multimodal behaviors to display and synchronizes the verbal and nonverbal behaviors of the agent. While other partners of the GVLEX project deal with expressive voice, our work focuses on expressive behaviors, especially on gestures.

The expressive gesture model is based at first on an existing virtual agent system, the GRETA system [1]. However, using a virtual agent framework for a

S. D´Mello et al. (Eds.): ACII 2011, Part II, LNCS 6975, pp. 224–231, 2011.

physical robot raises several issues to be addressed. Both agent systems, virtual and physical, have different degrees of freedom. Additionally, the robot is a physical entity with a body mass and physical joints which have a limit in movement speed. This is not the case of the virtual agent. Our proposed solution is to use the same representation language to control the virtual agent and the physical agent, here the robot Nao. This allows using the same algorithms for selecting and planning gestures, but different algorithms for creating the animation.

The primary goal of our research is to build an expressive robot able to display communicative gestures with different behavior qualities. In our model, the communicative gestures are ensured to be tightly tied with the speech uttered by the robot. Concerning the gestural expressivity, we have designed and implemented a set of quality dimensions such as the amplitude (SPC), fluidity (FLD), power (PWR) or speed of gestures (TMP) that has been previously developed for the virtual agent Greta [6]. Our model takes into account the physical characteristics of the robot.

2 Methodology

We have extended and developed the GRETA system to be used to control both, virtual and physical, agents. The existing behavior planner module of the GRETA system remains unchanged. On the other hand, a new behavior realizer module and a gestural database have been built to compute and realize the animation of the robot and of the virtual agent respectively. The detail of the developed system is described in the Section 4.

As a whole, the system calculates nonverbal behavior that the robot must show to communicate a text in a certain way. The selection and planning of gestures are based on information that enriches the input text. Once selected, the gestures are planned to be expressive and to synchronize with speech, then they are realized by the robot. To calculate their animation, gestures are transformed into key poses. Each key pose contains joint values of the robot and the timing of its movement. The animation module is script-based. That means the animation is specified and described with the multimodal representation markup language BML [11]. As the robot has some physical constraints, the scripts are calculated making it feasible for the robot.

Gestures of the robot are stored in a library of behaviors, called Lexicon, and described symbolically with an extension of the language BML [11]. These gestures are elaborated using gestural annotations extracted from a storytelling video corpus [14]. Each gesture in robot lexicon should be verified to make it executable for the robot (e.g. avoid collision or singular positions). When gestures are selected and realized, their expressivity is increased by considering a set of six parameters of gestural dimensions [6].

3 State of the Art

Several expressive robots are being developed. Behavior expressivity is often driven using puppeteer technique [13,22]. For example, Xing and co-authors propose to compute robot's expressive gestures by combining a set of primitive movements. The four movement primitives include: walking involving legs movement, swing-arm for keeping the balance in particular while walking, move-arm to reach a point in space and collision-avoid to avoid colliding with wires. A repertoire of gestures is built by combining primitives sequentially or additionally. However this approach is difficult to apply to humanoid robots with a serial motor-linkage structure.

Imitation is also used to drive a robot's expressive behaviors [3,8]. Hiraiwa et al [8] uses EMG signals extracted from a human to drive a robot's arm and hand gesture. The robot replicates the gestures of the human in a quite precise manner. This technique allows communication via the network. Accordingly, the robot can act as an avatar of the human. Another example is Kaspar [3]. Contrary to work aiming at simulating highly realistic human-like robot, the authors are looking for salient behaviors in communication. They define a set of minimal expressive parameters that ensures a rich human-robot interaction. The robot can imitate some of the human's behaviors. It is used in various projects related to developmental studies and interaction games.

In the domain of virtual agents, existing expressivity models either act as filters over an animation or modulate the gesture specification ahead of time. EMOTE implements the effort and shape components of the Laban Movement Analysis [2]. These parameters affect the wrist location of the humanoid. They act as a filter on the overall animation of the virtual humanoid. On the other hand, a model of nonverbal behavior expressivity has been defined that acts on the synthesis computation of a behavior [5]. It is based on perceptual studies conducted by Wallbott [21]. Among a large set of variables that are considered in the perceptual studies, six parameters [6] are retained and implemented in the Greta ECA system. Other works are based on motion capture to acquire the expressivity of behaviors during a physical action, a walk or a run [16].

Recent works have tendency to develop a common architecture to drive gestures of both virtual and physical agent systems. For instance, Salem et al. [20] develop the gesture engine of the virtual agent Max to control the humanoid robot ASIMO through a representation language, namely MURML. Nozawa et al. [18] use a single MPML program to generate pointing gestures for both animated character on 2D screen and humanoid robot in 3D space.

Similarly to the approaches of Salem and Nozawa, we have developed a common BML realizer to control the Greta agent and the NAO robot. Compared to other works, our system focus on implementing the gestural expressivity parameters [6] while taking into account the robot's physical constraints.

4 System Overview

The approach proposed in this thesis relies on the system of the conversational agent Greta following the architecture of SAIBA (cf. Figure 1). It consists of three separated modules [11]: (i) The first module, Intent Planning, defines communicative intents to be conveyed; (ii) The second, Behavior Planning, plans corresponding multimodal behaviors to be realized; (iii) and the third module, Behavior Realizer, synchronizes and realizes the planned behaviors. The results of the first module is the input of the second module through an interface described with a representation markup language, named FML (Function Markup Language). The output of the second module is encoded with another representation language, named BML [11]] and then sent to the third module. Both languages FML and BML are XML-based and do not refer to specific animation parameters of the agent (e.g. wrist joint).

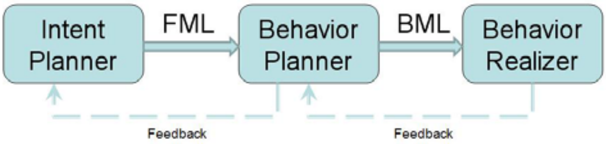


Fig. 1. SAIBA framework for multimodal behavior generation [11]

Aiming at being able to use the same system to control both agents (i.e. the virtual one and the physique one), however, the robot and the agent do not have the same behavior capacities (e.g. the robot can move its legs and torso but does not have facial expression and has very limited arm movements). Therefore the nonverbal behaviors to be displayed by the robot should be different from those of the virtual agent. For instance, the robot has only 2 hand configurations, open and closed; it cannot extend one finger only. Thus, to do a deictic gesture it can make use of its whole right arm to point at a target rather than using an extended index finger as done by the virtual agent. To control communicative behaviors of the robot and the virtual agent, while taking into account the physical constraint of both, two lexicons are taken into consideration, one for the robot and one for the agent. The Behavior Planning module of the GRETA framework remains the same. From the BML file outputted by the Behavior Planner, we instantiate the BML tags from either gestural repertoires. That is, given a set of intentions and emotions to convey, GRETA computes, through the Behavior Planning, the corresponding sequence of behaviors specified with BML. At the Behavior Realizer layer, some extensions are added to be able to generate animation specific to different embodiments (i.e. Nao and Greta). Firstly, the BML message received from Behavior Planner is interpreted and scheduled by a sub-layer called Animations Computation. This module is common for both agents. Then, an embodiment dependent sub-layer, namely Animation Production, generates and executes the animation corresponding to the specific implementation of agent. Figure 2 presents an overview of our system.

To ensure that both the robot and the virtual agent convey similar information, their gestural repertoires should have entries for the same list of communicative intentions. The elaboration of repertoires encompasses the notion of gestural family with variants proposed by Calbris [9]. Gestures from the same

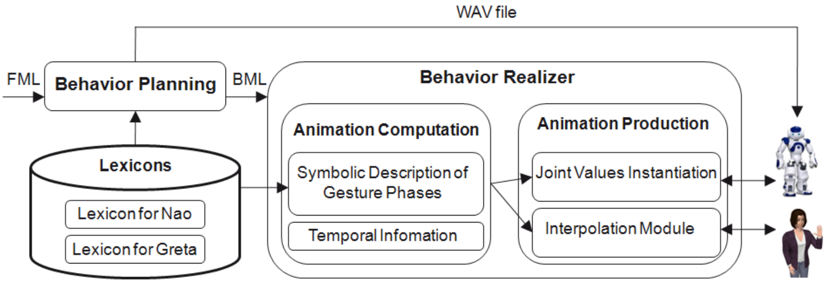


Fig. 2. An overview of the proposed system

family convey similar meanings but may differ in their shape (i.e. the element deictic exists in both lexicons; it corresponds to an extended finger or to an arm extension).

5 Gestural Lexicon

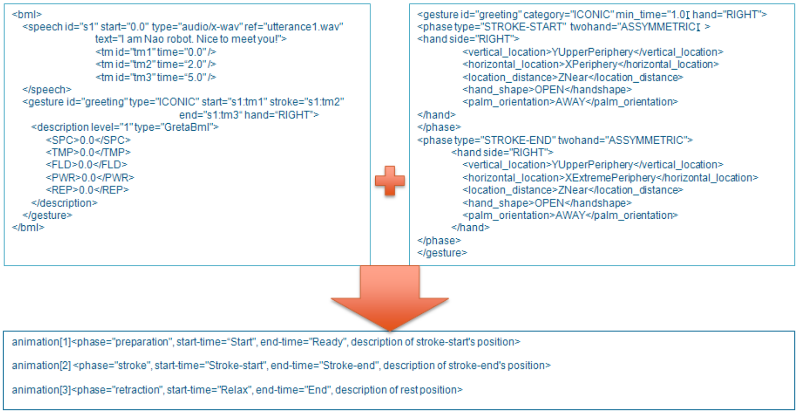
Symbolic gestures are stored in a repertoire of gestures (i.e. gestural lexicon) using an extension of the BML language. We rely on the description of gestures of McNeill [15], the gestural hierarchy of Kendon [10] and some notions from the HamNoSys system [19] to specify a gesture. As a result, a gestural action may be divided into several phases of wrist movement, in which the obligatory phase is called *stroke* transmitting the meaning of the gesture. The stroke may be preceded by a preparatory phase which puts the hand(s) of agent to the position ready for the stroke phase. After that it may be followed by a retraction phase that returns the hand(s) of agent to relax position or a position initialized by the next gesture.

In the lexicon, only the description of stroke phase is specified for each gesture. Other phases will be generated automatically by the system. A stroke phase is represented through a sequence of key poses, each of which is described with the information of hand shape, wrist position, palm orientation, etc.

The elaboration of gestures is based on gestural annotations extracted from a Storytelling Video Corpus [14]. All gestural lexicon are tested to guarantee its realizability on the robot.

6 Behavior Realizer

The main task of Behavior Realizer (BR) is to generate an animation of the agent (virtual or physical) from a BML message. This message contains descriptions of signals, their temporal information and values of expressivity parameters as illustrated in the Figure 3. The process is divided into two main stages. The first stage, called Animation Computation can be used in common for both agents while the second, Animation Production is specific to a given agent. This



```
<bml>
   <speech id="s1" start="0.0" type="audio/x-wav" ref="utterance1.wav"
                        text="I am Nao robot. Nice to meet you!">
                        <tm id="tm1" time="0.0" />
                        <tm id="tm2" time="2.0" />
                        <tm id="tm3" time="5.0" />
   </speech>
   <gesture id="greeting" type="ICONIC" start="s1:tm1" stroke="s1:tm2"
                                          end="s1:tm3" hand="RIGHT">
     <description level="1" type="GretaBml">
         <SPC>0.0</SPC>
         <TMP>0.0</TMP>
         <FLD>0.0</FLD>
         <PWR>0.0</PWR>
         <REP>0.0</REP>
     </description>
   </gesture>
</bml>
```

```
<gesture id="greeting" category="ICONIC" min_time="1.0I hand="RIGHT">
<phase type="STROKE-START"  twohand="ASSYMMETRICI >
<hand side="RIGHT">
            <vertical_location>YUpperPeriphery</vertical_location>
            <horizontal_location>XPeriphery</horizontal_location>
            <location_distance>ZNear</location_distance>
            <hand_shape>OPEN</handshape>
            <palm_orientation>AWAY</palm_orientation>
</hand>
</phase>
<phase type="STROKE-END" twohand="ASSYMMETRIC">
         <hand side="RIGHT">
            <vertical_location>YUpperPeriphery</vertical_location>
            <horizontal_location>XExtremePeriphery</horizontal_location>
            <location_distance>ZNear</location_distance>
            <hand_shape>OPEN</handshape>
            <palm_orientation>AWAY</palm_orientation>
         </hand>
</phase>
</gesture>
```

```
animation[1]<phase="preparation", start-time="Start", end-time="Ready", description of stroke-start's position>

animation[2] <phase="stroke", start-time="Stroke-start", end-time="Stroke-end", description of stroke-end's position>

animation[3]<phase="retraction", start-time="Relax", end-time="End", description of rest position>
```

Fig. 3. An example of data processing in Behavior Realizer

architecture is similar to the system proposed by the Heloir et al. [7]. However our solution is different that it aims at different embodiments [17].

6.1 Synchronization of Gestures with Speech

The synchronization is obtained by adapting gestures to the speech structure. The temporal information of gestures in BML tags are relative to the speech by time markers (see Figure 3). The time of each gestural phase is indicated with several sync points (start, ready, stroke-start, stroke, stroke-end, relax, end). Following an observation of McNeill [15], the stroke phase coincides or precedes emphasized words of the speech. Hence the timing of the stroke phase should be indicated in the BML description. Meanwhile the timing of other phases is supposed to be based on the timing of the stroke phase.

6.2 Animation Computation

This module analyses a BML message received from Behavior Planner and loads corresponding gestures from the lexicon. Based on available and necessary time, the module checks if a gesture is executable. Then it calculates symbolic values and timing for each gesture phase while taking into account gestural expressivity parameters (e.g. the duration of gestural stroke phase is decreased when the temporal extension (TMP) is increased and vice-versa).

6.3 Animation Production

In this stage, the system generates the animation by instantiating the symbolic description of the planned gesture phases into joint values thanks to the Joint Values Instantiation module. One symbolic position will be translated into concrete values of six robot joints (ShoulderPitch, ShoulderRoll, ElbowYaw, ElbowRoll, WristYaw, Hand) [17]. Then the animation is obtained by interpolating between joint values with robot built-in proprietary procedures [4].

7 Conclusion and Future Work

We have presented an expressive gesture model that is designed and implemented for the humanoid robot Nao. The model focuses not only on the coordination of gestures and speech but also on the signification and the expressivity of gestures conveyed by the robot. While the gestural signication is studied carefully when elaborating a repertoire of robot gestures (i.e. lexicon), the gestural expressivity is increased by adding the gestural dimension parameters specified by the GRETA system. A procedure creating a gestural lexicon overcoming physical constraints of the robot has been defined.

In the future, we propose to implement all of the expressivity parameters and to valid the model through perceptive evaluations. As well the implementation of gesture animation and expressivity should be evaluated. An objective evaluation will be set to measure the capability of the implementation. A subjective evaluation will be made to test how expressive the gesture animation is perceived on the robot when reading a story.

Acknowledgment. This work has been funded by the GVLEX project (http://www.gvlex.com).

References

1. Pelachaud, C., Gelin, R., Martin, J.-C., Le, Q.A.: Expressive Gestures Displayed by a Humanoid Robot during a Storytelling Application. In: New Frontiers in Human-Robot Interaction (AISB), Leicester, GB (2010)
2. Chi, D., Costa, M., Zhao, L., Badler, N.: The EMOTE model for effort and shape. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 173–182 (2000)
3. Dautenhahn, K., Nehaniv, C., Walters, M., Robins, B., Kose-Bagci, H., Mirza, N., Blow, M.: Kaspar–a minimally expressive humanoid robot for human–robot interaction research. Applied Bionics and Biomechanics 6(3), 369–397 (2009)
4. Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., Maisonnier, B.: Mechatronic design of NAO humanoid. In: Robotics and Automation, ICRA 2009, pp. 769–774. IEEE, Los Alamitos (2009)
5. Hartmann, B., Mancini, M., Pelachaud, C.: Towards affective agent action: Modelling expressive ECA gestures. In: International Conference on Intelligent User Interfaces-Workshop on Affective Interaction, San Diego, CA (2005)
6. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS (LNAI), vol. 3881, pp. 188–199. Springer, Heidelberg (2006)
7. Heloir, A., Kipp, M.: EMBR – A realtime animation engine for interactive embodied agents. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 393–404. Springer, Heidelberg (2009)

8. Hiraiwa, A., Hayashi, K., Manabe, H., Sugimura, T.: Life size humanoid robot that reproduces gestures as a communication terminal: appearance considerations. In: Computational Intelligence in Robotics and Automation, vol. 1, pp. 207–210. IEEE, Los Alamitos (2003)
9. Iverson, J., Goldin-Meadow, S.: Why people gesture when they speak. Nature 396(6708), 228–228 (1998)
10. Kendon, A.: Gesture: Visible action as utterance. Cambridge University Press, Cambridge (2004)
11. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjálmsson, H.: Towards a common framework for multimodal generation: The behavior markup language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 205–217. Springer, Heidelberg (2006)
12. Krauss, R.: Why do we gesture when we speak? Current Directions in Psychological Science 7(2), 54–60 (1998)
13. Lee, J., Toscano, R., Stiehl, W., Breazeal, C.: The design of a semi-autonomous robot avatar for family communication and education. In: Robot and Human Interactive Communication (ROMAN 2008), pp. 166–173. IEEE, Los Alamitos (2008)
14. Martin, J.C.: The contact video corpus (2009)
15. McNeill, D.: Hand and mind: What gestures reveal about thought. University of Chicago Press, Chicago (1992)
16. Neff, M., Fiume, E.: Methods for exploring expressive stance. Graphical Models 68(2), 133–157 (2006)
17. Niewiadomski, R., Bevacqua, E., Le, Q., Obaid, M., Looser, J., Pelachaud, C.: Cross-media agent platform. In: 16th International Conference on 3D Web Technology (2011)
18. Nozawa, Y., Dohi, H., Iba, H., Ishizuka, M.: Humanoid robot presentation controlled by multimodal presentation markup language mpml. In: Robot and Human Interactive Communication (ROMAN 2004), pp. 153–158. IEEE, Los Alamitos (2004)
19. Prillwitz, S.: HamNoSys Version 2.0: Hamburg notation system for sign languages: An introductory guide. Signum (1989)
20. Salem, M., Kopp, S., Wachsmuth, I., Joublin, F.: Generating robot gesture using a virtual agent framework. In: Intelligent Robots and Systems (IROS 2010), pp. 3592–3597. IEEE, Los Alamitos (2010)
21. Wallbott, H.: Bodily expression of emotion. European Journal of Social Psychology 28(6), 879–896 (1998)
22. Xing, S., Chen, I.: Design expressive behaviors for robotic puppet. In: Control, Automation, Robotics and Vision (ICARCV 2002), vol. 1, pp. 378–383. IEEE, Los Alamitos (2002)