Expressive Gestures Displayed by a Humanoid Robot during a Storytelling Application

Catherine Pelachaud[1] and **Rodolphe Gelin**[2] and **Jean-Claude Martin**[3] and **Quoc Anh Le**[4]

Abstract. Our purpose is to have the humanoid robot, NAO, to read a story aloud through expressive verbal and nonverbal behaviors. These behaviors are linked to the story being told and the emotions to be conveyed. To this aim, we are gathering a repertoire of gestures and body-postures for the robot using a video corpus of story-tellers. The robot's behavior is controlled by the Greta platform that drives agents to communicate through verbal and nonverbal means. In this paper we are presenting our research methodology and overall framework.

1 Introduction

Expressivity is an important factor in communication, be human-human, human-virtual agent or human-robot communication. It is linked to emotional states [30]; it also reflects some idiosyncratic features [11]; it has several communicative functions such as calling for attention or showing contrast [6]. Expressivity can be conveyed multimodally, through the voice and the body. Communication involves much more than word. Prosody, gesture, facial expression, gaze, body posture, all participate to conveying a message by a sender (the speaker) to an addressee (the listener). Gesture convey complementary features compared to speech (see Kendon's continuum described by [24]). Furthermore they have been observed to play a role in the expression of emotion [30, 16]. Our idea is to endow a humanoid robot with human communicative capabilities.

Nao is a humanoid robot able to display body movement and gesture. It can talk and interact with humans. In our project, our aim is to have Nao reads expressively a story. This objective implies to enhance the expressivity of the speech synthesizer and the behavior animation technology. When listening to a story being told without expressivity, it is difficult to perceive the various peripeteia of the story and to personalize the different characters of the story. Moreover one gets disinterested rapidly; it is difficult to maintain one's attention. It is even more so when the story is being told by a humanoid remaining still, not displaying any body movements. To have the robot reads expressively, we foresee two main research directions: extract information from the story to be said and define the nonverbal (acoustic and visual) behaviors to said it aloud. Obtaining a thorough analysis of the text of the story is important. One needs to interpret its content: to extract information on its structure, its various semantic and pragmatic elements as well as its emotional content [3, 28]. Detailed information from the story is crucial to drive expressive speech and behavior synthesizer. This work is part of a French national project named GV-Lex (Gesture and Voice for an Expressive Reading). Within the GV-Lex project our aim is not only to compute an expressive voice but also an expressive animation for a given application: reading tales to children.

Our idea for driving the behavior of the robot is to apply the multimodal behavior generation model implemented in the Greta platform [26]. To this aim we have enhanced the repertoire of gestures of the robot and of the agent. Since they do not have the same motion capacities we are developing a gesture representation schema based on the notion of gesture variant [4]. The resulting animation of the robot and of the virtual agent may not be identical but would convey the same meaning. Thus, at first, we gather information of the type of gestures and body postures that are displayed when telling a story. This is done by conducting a careful analysis of a video corpus of story-tellers. Particular attention has been put on the emotional information body posture conveys. Body posture is defined as a combination of head direction, arm shape and lower body position. The analysis also provides information of the communicative gestures that are used to convey a given intention. On a second stage, we are using the Greta platform to compute the robot's animation. Given an emotional state and communicative intention to transfer, the Greta platform computes which non-verbal behaviors to use and with which expressivity. It is SAIBA compliant [29] and uses two representation languages, one at the intention level FML [13] and one at the behavior level BML [29]. A repertoire of behaviors, specified in BML, contains the behaviors the robot can use. This repertoire is constituted of pairs where one element is the description of the behavior and the second element the meaning attached to it.

In this paper we focus on the nonverbal behavior of the robot. We present our preliminary result and explain our research methodology. After an overview of the state of the art, we present Nao, the humanoid robot we use. We argument how Nao can be accepted as a story teller and which capacities it requires. We then turn our attention to the behavior engine we use. We follow by describing the video corpus of story-tellers and by explaining how multimodal behaviors of actors were recorded when telling the same story. Finally we present a gesture specification for both the robot and the virtual agent. A conclusion ends the paper.

2 State of the art

Several expressive robots are being developed. Behavior expressivity is often driven using puppeteer technique [21, 31]. For example Xing and co-authors propose to compute robot's expressive gestures by combining a set of primitive movements. The four movement primi-

[1] CNRS, LTCI Telecom ParisTech, France, email: catherine.pelachaud@telecom-paristech.fr

[2] Aldebaran, France email: rgelin@aldebaran-robotics.com

[3] CNRS, LIMSI, email: martin@limsi.fr

[4] CNRS, LTCI Telecom ParisTech, France, email: quoc@telecom-paristech.fr

tives are: walk 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.

Imitation is also used to drive a robot's expressive behaviors [14, 7]. Hiraiwa et al [14] 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. The robot can act as an avatar of the human. Another example is Kaspar [7]. 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 acts as filters over an animation or modulates the gesture specification ahead of time. Emote implements the effort and shape components of the Laban Movement Analysis. 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 [12]. It is based on perceptual studies conducted by Wallbott [30]. Among the large set of variables that are considered in the perceptual studies, six parameters [12] were retained and implemented in the Greta ECA system [27]. Other works are based on motion capture to acquire the expressivity of behaviors during a physical action, a walk or a run [25].

3 Acceptability of domestic humanoid robots

Since 2005, Aldebaran Robotics produces Nao, a 57cm high humanoid robot fully motorized. At the end of 2009, more than 400 Naos have been sold all over the world. Today, the users of Nao are mainly researchers; however Aldebaran is convinced that humanoid robot could become the best robotic assistant at home. There are several objective reasons that motivate our choice of create Nao as a robot with a humanoid appearance and communicative capabilities. Since domestic environments are designed for human beings, the human shape is the most appropriate to intervene in these environments: catching objects on a table or in a closet requires arms, hands and vertical posture; bipedal locomotion is not only necessary to climb stairs but is also more convenient to move in cluttered environments. Among the best researchers are working on these necessary functions [20] [8] which are still only available in laboratories. A lot of work has to be done before the robot can be used outside a controlled environment, into the wild.

From a communication point of view, the humanoid robot is much more attractive than nonhuman-like machines. With its body motions (Nao does not have animated mouth, eyes and eyebrows), the humanoid robot is able to catch attention and often affection of people. By imitating or caricaturing expressive motions of humans, the humanoid robot creates a complicity with its human interlocutors [10]. Movements of a humanoid robot will be considered as communicative gestures if they are natural, encompass dynamic quality and convey appropriate meaning that its human interlocutors can understand. Considering the number of degrees freedom (DOF) of a humanoid robot (Nao has 25 DOF), it appears that programming the motion of each axis independently, even in Cartesian mode, is very time consuming when subtle expressive motions are expected.

4 Expressive gesture model in Nao

To ensure Nao can produce communicative gestures, Aldebaran developed a very powerful tool for programming motions on Nao, Choregraphe. This graphical tool has been presented in [9]. It allows one to synchronize the motion of Nao limbs intuitively. Working with actors, ethnologists and character animators coming from movie industry, Aldebaran has created a library of expressive gestures illustrating different concepts: objects (ball, weapon, pencil...), situations (I am lost, I don't understand, my batteries are empty) or even emotions (levels of happiness, fear or sadness). Accompanied with relevant sounds, these gestures give to Nao a real presence. Today more than 200 meaningful animations are available in the library of Nao.

These animations can be called explicitly within a program. For instance, the robot is looking for an object. When it sees this object, it triggers the "happy dance" animation. This very procedural use of expressive gestures is interesting but not so efficient. Each animation is associated with a semantic meaning. Each of them can be represented by a couple (meaning, intensity) where the first element of the couple, meaning, can take the value of an emotional state (eg happiness, sadness, fear) and the second element represents the intensity (low, normal, high) of a given meaning. According to this very simple model, a program running on Nao can call a function "Expressive_motion" with two arguments (emotion, intensity). So far, this function will trigger randomly one of the expressive animations in the corresponding category.

But before the beginning of the project, Aldebaran Robotics developed a software, called "Narrateur" [2] that proposes to read texts accompanied with expressive animations. If the robot only uses a Text To Speech (TTS) software to read texts, it does not offer much more than a computer. In order to show the advantage of a robot as text reader expressive gestures illustrating the text are necessary. The off-line part of Narrateur proposes a simple interface allowing the inclusion, in the text to be read, of beacons that triggered given animations. A smiley at the end of the sentence will be interpreted by an "happy dance" animation for instance. The input text is parsed at run-time. The sentences are sent to the TTS and the beacons are sent to the animation generator of Nao. When "Narrateur" messages are read by Nao, people are generally seduced by the little robot. This kind of simple functions could convince people to have a humanoid robot at home for fun, in a first time, waiting for a more efficient domestic robot in a diversity of application domains.

If Narrateur is a first efficient step towards the development of an expressive robot, its principle to describe animations is at a too macroscopic level. Its main drawback is the difficulty to synchronize precisely the text and the gestures. For instance, the robot will not be able to synchronise gesture and intonation (on the words "little" and the "big" of the sentence "the little girl meets the big bad wolf". Precised synchronisation of gesture and intonation are important to convey expressivity.

5 Behavior Engine

We aim to use an existing behavior engine to compute the nonverbal behavior of the robot [27]. The Greta platform is SAIBA compliant [29]. It is composed of three main modules:

Intent Planner determines the agent's communicative intentions and emotional states

Behavior Planner decides and schedules which verbal and nonverbal behaviors to use to convey these intentions and emotional states

Behavior Realizer computes the animation to be visualized by the agent

A gestuary is a repository of nonverbal behaviors. Each element of this repository is a pair where one element corresponds to the meaning carried out by the behavior and the second one to the shape description of the behavior. Communicative behaviors are polysemic. They can be attached to different meanings. Identically, a given meaning can usually be expressed by several behaviors. Two main languages have been defined to encode dataflow over these three modules. FML stands for Function Markup Language. It encodes the communicative intentions and emotional states that are outputted by the Intent Planner. BML, Behavior Markup Language, links the behavior planner to the behavior realizer. It specifies the behaviors to be displayed. BML specification is independent of the animation parameters of the agent. Our aim is to use the Greta platform to control a virtual agent, Greta, and a physical agent, Nao. To this aim we are currently developing a gestuary for Nao. The gestuary is being created using information extracted from a corpus analysis that is explained in the next section. The animation of both agents type is driven by FML and BML tags. In a past study, we have used a similar approach with the Sony Aibo robot [1]. One gestuary of backchannel signals has been defined for the robot and one for the virtual agent. The animation of both agents was driven from the same architecture.

6 A multimodal video corpus of storytelling

Corpus-based models of multimodal behaviors are useful for the design of expressive agents and other multimodal interfaces [17]. Designers of embodied agents and humanoid robots might inspire from a variety of complementary knowledge sources in order to specify naturalistic and context specific multimodal expressions: literature, corpora, and results of perception tests. For example, motion capture corpora can be useful for inspiring the design of postural expressions by humanoid robots [19]. As behaviors of individual human subjects are recorded in video corpora using such a technique, individual gesture profiles can be computed [15]. Such individual gesture profile can be useful for designing storytelling agents and robots with individual styles that can be adapted to the audience at hand. Several experimental and corpus based studies about emotional expression observed discriminative features of emotions categories or dimensions in the global posture, the movement quality [30] and basic gestural form features [16].

The ContAct video corpus [22] was recorded in order to collect data on how storytellers use expressive gestures in a specific storytelling context. One French tale was selected according to several criteria. The duration of the tale was supposed to be long enough to enable the expression of a wide variety of emotional states and gestures (but also not too long so that the audience listening to the story read by a virtual agent or a robot would not get distracted). We selected a French tale called ”Three Pieces of Night”. It is a thousand words long and relates both positive and negative events. Six actors from a local amateur troupe were videotaped while telling this story so that we would be able to study a variety of styles when relating the same story. Two digital cameras were used (front view and side-view) as postural expression of emotion is expected to be three dimensional [18]. Actors had received the script of the story a few days before the recording session. The text was also displayed on the wall during the session so that subjects could read it from time to time. Each subject was recorded twice so that the most expressive session would be kept for analysis. The collected corpus contains 80 mn of videos (average 7,5 mn for telling the story).

Gestures observed in the corpus will be annotated to inspire the specification of expressive gestures to be used by the Greta agent and the Nao robot. A scheme that has been already been applied successfully in a similar copy-synthesis approach will be adapted [23].

7 Gestures specification for virtual and physical agents

A difficulty arises from the differences in the agents’ body specification, if it is virtual or physics. While Nao can not show facial expressions, Greta does not move its lower body. Nao can also use very few hand shapes but it can keep its equilibrium. Greta can display any hand shapes but has no sense of gravity. These variations result in different animations types.

Even though the bodies of the virtual and physical agents are different, if they are controlled by the same FML inputs, their animation should convey the same meanings. That is the sets of BML tags outputted for Nao and for Greta should be similar in meanings (but not necessarily in behaviors) (see Figure 1). To ensure to maintain the same meaning we are expanding the Gestuary for each agent as follows: first we enhance the gestuary of both agents to have similar behaviors as possible. We are also using the notion of gesture variants that was first used by N. Ech Chafai [5]. Gesture variant was introduced by Geneviève Calbris, a French semiologist that studied communicative gestures [4]. Variant of a gesture encompasses a family of gestures that shares the same meaning (eg to negate something) and a core signal (eg vertical flat hand toward the other). Gestures within a family may differ along the non-core signals they use (eg frowning). Thus in the Greta-Gestuary and the Nao-Gestuary, gesture variant shares similar meaning and signal-core. However they differ along the other signals (eg use of frown vs red leds). When the signal-core of a given gesture can be displayed with Greta but is not possible at all in Nao, the gesture entry appears only in Greta’s lexicon. Other entries for the same meaning are nevertheless included in Nao’s lexicon, so that the meaning can be expressed using another gesture.

8 Conclusion

In this paper we have presented preliminary work we are conducting to develop an expressive robot capable of reading stories aloud. The expressive behavior lexicon for the robot is based on a video corpus analysis and is described using BML. We are driving the robot’s animation using the Greta platform. We need to carry out an evaluation to ensure the robot Nao conveys similar intentions and emotional states as the virtual agent Greta, given a same FML input. We will also perform user tests to validate the robot can read expressively and can maintain human users’ interest and engagement. Such perception tests will be conducted by comparing the perception of the story told either by a human (video corpus), the robot with / without expressive gestures or the expressive agent. Both objective (recall of the story and of its emotional content) and subjective measures (preference, perceived expressiveness) will be investigated.

ACKNOWLEDGEMENTS

This work has been partially funded by the French ANR project GV-Lex (ANR-08-CORD-024 (http://www.gvlex.com/).

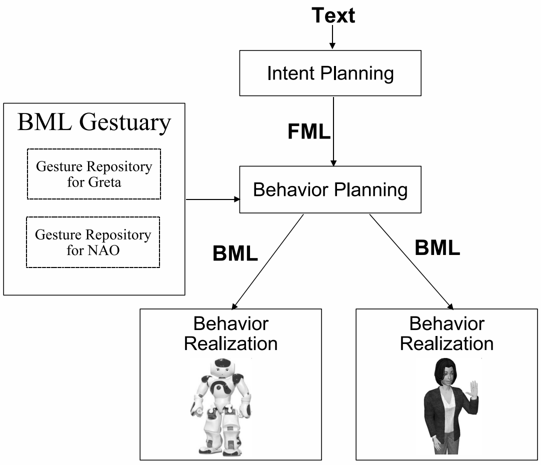

Figure 1. Greta platform driving Nao and Greta

REFERENCES

[1] S. AL Moubayed, M. Baklouti, M. Chetouani, T. Dutoit, A. Mahdhaoui, J-C Martin, S. Ondas, C. Pelachaud, J. Urbain, and M. Yilmaz, ‘Generating robot/agent backchannels during a storytelling experiment’, in *IEEE International Conference on Robotics and Automation (ICRA’09), Japan, 2009.*, (2009).

[2] Maisonnier B and Monceaux J. Système et procédé pour générer des comportements contextuels d’un robot mobile. French Patent 66058 FR, 07 2009. Deposit number : 09 54837.

[3] S. Le Beux, A. Rilliard, and C. dAlessandro, ‘Calliphony: A real-time intonation controller for expressive speech synthesis’, in *6th ISCA Workshop on Speech Synthesis (SSW-6)*, Bonn, Germany, (August 22-24 2007).

[4] G. Calbris, *The semiotics of French gestures*, University Press, Bloomington: Indiana, 1990.

[5] N.E. Chafai, C. Pelachaud, and D. Pelé, ‘Towards the specification of an ECA with variants of gestures’, in *Proceedings of Intelligent Virtual Agents, IVA’07*, Paris, (September 2007).

[6] N.E. Chafai, C. Pelachaud, and D. Pelé, ‘A case study of gesture expressivity breaks’, *International Journal of Language Resources and Evaluation, Special issue on Multimodal Corpora for Modelling Human Multimodal Behavior*, (2008).

[7] Kerstin Dautenhahn, Chrystopher L. Nehani, Michael L. Walters, Ben Robins, Hatice Kose-Bagci, N. Assif Mirza, and Mike Blow, ‘KASPAR A minimally expressive humanoid robot for human-robot interaction research’, *Applied Bionics and Biomechanics*, (submitted).

[8] Dimitar Dimitrov, Pierre-Brice Wieber, Olivier Stasse, Joachim Ferreau, and Holger Diedam, ‘An Optimized Linear Model Predictive Control Solver for Online Walking Motion Generation’, in *IEEE International Conference on Robotics & Automation*, Kobe Japon, (2009).

[9] E Pot et al., ‘Choregraphe: Graphical tool for humanoid robot programming’, in *IEEEROMAN*, (2009).

[10] J. Monceaux et al., ‘First steps in emotional expression of the humanoid robot nao’, in *IEEE-RAS Humanoids*, pp. 331–336, Paris France, (2009).

[11] P.E. Gallaher, ‘Individual differences in nonverbal behavior: Dimensions of style’, *Journal of Personality and Social Psychology*, **63**(1), 133–145, (1992).

[12] Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud, ‘Implementing expressive gesture synthesis for embodied conversational agents.’, in *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Berder Island*, pp. 188–199, (2005).

[13] D. Heylen, S. Kopp, S. Marsella, C. Pelachaud, and H. Vilhjalmsson, eds. *Why Conversational Agents do what they do? Functional Representations for Generating Conversational Agent Behavior The First Functional Markup Language Workshop*, 2008. The Seventh International Conference on Autonomous Agents and Multiagent Systems Estoril, Portugal.

[14] Akira Hiraiwa, Kouk Hayashi, Hiroyula Manabe, and Toshiaki Sugimura, ‘Life size humanoid robot that reproduces gestures as a communication terminal: Appearance considerations’, in *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Kobe, Japan, (July 16-20 2003).

[15] M. Kipp, *Gesture Generation by Imitation. From Human Behavior to Computer Character Animation*, Boca Raton, Dissertation.com, Florida, 2004.

[16] M. Kipp and J.-C. Martin, ‘Gesture and emotion: Can basic gestural form features discriminate emotions?’, in *International Conference on Affective Computing and Intelligent Interaction (ACII-09)*. IEEE Press, (2009).

[17] M. Kipp, J.C. Martin, P. Paggio, and D. Heylen, *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*, volume Vol. 5509 of *Lecture Notes on Artificial Intelligence, LNAI 5509*, Springer, 2009.

[18] A. Kleinsmith and N. Bianchi-Berthouze, ‘Recognizing affective dimensions from body posture’, in *2nd International Conference on Affective Computing and Intelligent Interaction (ACII 2007)*, eds., A. Paiva, R. Prada, and R. Picard, pp. 48–58, Lisbon, Portugal, (2007). Springer, LNCS, vol. 4738.

[19] A. Kleinsmith, I. Rebai, N. Berthouze, and J.-C. Martin, ‘Postural expressions of emotion in motion captured database and in a humanoid robot’, in *International Workshop on Affective-aware Virtual Agents and Social Robots (AFFINE’09) held during the ICMI-MLMI’09 conference*, eds., G. Castellano, J.C. Martin, J. Murray, K. Karpouzis, and C. Peters, Boston, USA, (2009). ACM 978-1-60558-692-2-1/09/11.

[20] Marin-Urias L, Akin Sisbot E, Pandey A, Tadakuma R, and Alami R, ‘Towards shared attention through geometric reasoning for human robot interaction’, in *demonstration paper, ICMI MLMI*, Cambridge MA USA, (2009).

[21] Jun Ki Lee, R.L. Toscano, W.D. Stiehl, and C. Breazeal, ‘The design of a semi-autonomous robot avatar for family communication and education’, in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pp. 166–173, (Aug. 2008).

[22] J.-C. Martin. The contact video corpus, 2009.

[23] J.-C. Martin, R. Niewiadomski, L. Devillers, S. Buisine, and C. Pelachaud, ‘Multimodal complex emotions: Gesture expressivity and blended facial expressions’, *Special issue of the Journal of Humanoid Robotics on ”Achieving Human-like Qualities in Interactive Virtual and Physical Humanoids”. Eds: C. Pelachaud, L. Canamero.*, **3**(3), 269–291, (2006).

[24] D. McNeill, *Gesture and Thought*, University of Chicago Oress, Chicago, 2005.

[25] M. Neff and E. Fiume, ‘Methods for exploring expressive stance’, in *Proceedings of ACM, SIGGRAPH/EUROGRAPHICS Symposium of Computer Animation*, pp. 49–58, Grenoble, (August 2004).

[26] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud, ‘Greta: an interactive expressive ECA system’, in *Demo session at AAMAS’09*, Budapest, (May 2009).

[27] C. Pelachaud, ‘Modelling multimodal expression of emotion in a virtual agent’, *Philosophical Transactions of Royal Society B Biological Science*, **B 2009 364**, 3539–3548, (2009).

[28] Sophie Rosset, Delphine Tribout, and Lori Lamel, ‘Multi-level information and automatic dialog act detection in human-human spoken dialogs’, *Speech Communication*, **50**(1), (2008).

[29] H. Vilhjalmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. van der Werf, ‘The behavior markup language: Recent developments and challenges’, in *Intelligent Virtual Agents, IVA’07*, Paris, (September 2007).

[30] H.G. Wallbott, ‘Bodily expression of emotion’, *European Journal of Social Psychology*, **28**, 879–896, (1998).

[31] Shusong Xing and I-Ming Chen, ‘Design expressive behaviors for robotic puppet’, in *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on*, volume 1, pp. 378–383 vol.1, (Dec. 2002).