From Speech to Emotional Interaction: EmotiRob Project

Marc Le Tallec[1], Sébastien Saint-Aimé[2], Céline Jost[2], Jeanne Villaneau[2], Jean-Yves Antoine[1], Sabine Letellier-Zarshenas[2], Brigitte Le-Pévédic[2], and Dominique Duhaut[2]

[1] Computer sciences laboratory, University François-Rabelais, Tours, France
{marc.letallec,jean-yves.antoine}@univ-tours.fr
[2] Valoria laboratory, University of Bretagne Sud, Vannes, France
{sebastien.saint-aime,celine.jost,jeanne.villaneau,sabine.letellier,
brigitte.le-pevedic,dominique.duhaut}@univ-ubs.fr

Abstract. This article presents research work done in the domain of nonverbal emotional interaction for the EmotiRob project. It is a component of the MAPH project, the objective of which is to give comfort to vulnerable children and/or those undergoing long-term hospitalisation through the help of an emotional robot companion. It is important to note that we are not trying to reproduce human emotion and behavior, but trying to make a robot emotionally expressive. This paper will present the different hypotheses we have used from understanding to emotional reaction. We begin the article with a presentation of the MAPH and EmotiRob project. Then, we quickly describe the speech undestanding system, the iGrace computational model of emotions and integration of dynamics behavior. We conclude with a description of the architecture of Emi, as well as improvements to be made to its next generation.

Keywords: emotion, interaction, companion robot, dynamic behavior.

1 Introduction

Currently, research in robotics focuses on cooperative systems for performing complex tasks with humans. Another new challenge is establishing systems which provide enrichment behaviour through their interaction with humans. Research in psychology has demonstrated that facial expressions play an essential role in coordinating human conversation [1] and is a key modality of human communication. As social robots are very limited in their mechanical and intellectual capacities, they are mostly used for human amusement and leisure purposes. Robotherapy, a field in robotics, tries to apply the prinicples of social robotics to improve the psychological and physiological state of people who are ill, marginalized, or suffering from physical or mental disabilities. Within this context, the robots seem able to play a role of guidance and enlightenment, which requires providing them with as many communication capacities as possible.

We, therefore, began experiments [2] using the Paro robots to see whether or not reaction/interaction with robots is dependent on culutral contexts. These experiments showed us the two main directions which would forward our work. The first one deals with mechanical problems. The robot should be very light, easy to pick up and handle, at least easier than Paro. It should also have a great deal of autonomy. The second direction is toward changing man-machine interaction: the psychological comfort that the robot can provide is related to the quality of the emotional tie that the child has with it.

2   MAPH and EmotiRob Project

The objective of the MAPH project is to design an antonomous stuffed robot, which could bring some comfort to vulnerable children (eg. children enduring long hospital stay). However, it is essential to avoid designing a robot that is too complex and too voluminous.
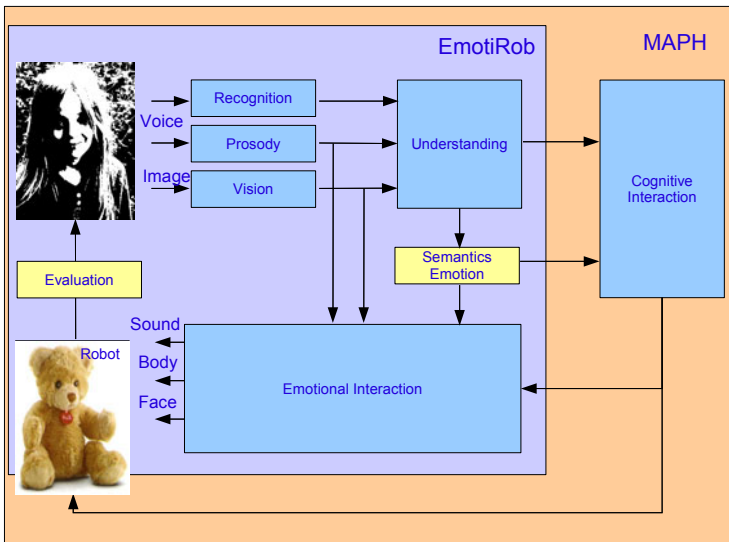

Fig. 1. Synoptic of MAPH project, including EmotiRob sub-project

Figure 1 shows our synopsis, giving the different modules which will forward good development of the MAPH and EmotiRob project:

- In our entry understanding module is the information of processing prosody, video, as well as voice recognition. These factors will enable us to gather the utterances of the child, as well as his emotional state.
- Once the entries are processed, this information will be forwarded to the emotional interaction module via a semantics emotional structure.

- A portion of the information handled by the entry module will allow the cognitive interaction module to determine an appropriate reaction to the behavior and discourse of the child.
- The output will transcribe the interaction and emotion of the robot through small sounds, body movements and facial expressions.

3   Understanding System and Detection of Emotion

Our system aims at detecting emotions conveyed in words used by children by combining prosodic and linguistic clues. Nevertheless, the subject of our current work is the detection of emotions from the propositional content of only the words used by children, by using the semantic structure of the utterances.

3.1   Spoken Language Understanding (SLU)

The SLU, which is used, is based on logical formalisms and achieves an incremental deep parsing [3]. It provides a logical formula to represent the meaning of the word list that Automatic Speech Recognition provides to it as input. The vocabulary known by the system as a source langage contains about 8000 lemmas selected from the lexical Manulex1[1] and Novlex2[2] bases. We have restricted the concepts of the target language by using Bassano's studies related to the development of child language [4]. SLU carries out a projection from the source language into Bassano's vocabulary information. Even if that domain is larger than most domains where SLU is used, it is not as large as the one used for the Emotirob project.

Thus, to adapt the system to our objective, we had to build an ontology from the set of application concepts. More precisely, the Bassano vocabulary included many verbs, some of which had extended and even polysemic meanings. To specify the possible uses of these verbs, a part of the ontology [5] is based on a linguistic corporus study related to fairy tales.

The parsing is split into three main steps: the first step is a chunking [6] which segments a sentence into minimal syntactic and semantic groups. The second step builds semantic relations between the resulting chunks and the third is a contextual interpretation. The second and third steps use a semantic knowledge of the application domain.

3.2   Emologus

In the Emologus system [7], the detection of emotions relies on a major principle: the emotion conveyed by an utterance is compositional. It depends on the emotion of every individual word, as well as the semantic relations characterized by the SLU system. More precisely, simple lexical words have an intrinsic emotional

---

[1] http://leadserv.u-bourgogne.fr/bases/manulex/manulexbase/indexFR.htm
[2] http://www2.mshs.univ-poitiers.fr/novlex/

value, while verbal and adjectival predicates act as a function on the emotional values of their arguments.

The first requirement is understanding the type of emotion that children associate to lexicon words. We used the standard emotional lexicon of Syssau and Monnier [8], who have compiled the evaluations provided by young children for French language words. We have completed the standard lexicon with the evaluation provided by children of 80 new words extracted from the Bassano lexicon.

To complete the characterization of our lexicon, an emotional predicate was assigned to every verb or adjective of our application lexicon through an agreement procedure among five adult experts. In fact, every expert proposed one or at most two definitions for every predicate. Then, agreement was sought among these proposals. It is interesting to note that a complete agreement was finally able to be reached.

We have tested Emologus on a corpus, composed of about 170 sentences which make up twenty short stories. We only have out-of-context sentences, but the results for these are encouraging. For out-of-context sentences, we show that it is possible to find the present emotion from linguistic clues in a sentence in 90% of cases. A very positive fact is that we never find an opposite emotion.

4   iGrace – Computational Model of Emotions

Before beginning our project, we did two experimental studies. The first experiment [2] was carried out using the Paro robot to verify if reaction/interaction with robots depended on cultural context. This experiment highlighted the fact that there could be mechanical problems linked to weight and autonomy, as well as interaction problems due to the lack of emotions of the robot.

The second experiment [9] was to help us reconcile the restriction of a light, autonomous robot with understanding expression capacities. Evaluators had to select the faces that seemed to best express primary emotions among a list of 16 faces. It was one of the simplest faces that obtained the best results. With only 6 degrees of freedom [10], it was possible to obtain a very satisfying primary emotion recognition rate.

With this information, we began working on the conception of our emotional interaction model. iGrace [11], based on the GRACE emotional model [12] that we designed, will help us to attain our research objectives. It is composed of 3 main models (descriptions of which will be given in the following subsections) which will be able to process the received information:

- The "Input" Module represents the interface for communication and data exchange between the understanding module and emotional interaction module.
- The "Emotional interaction" Module will carry out the processing necessary for interaction in six steps:
  1. Extraction, from list L_1, of emotional experiences linked to the personality of the robot
  2. Extraction, from list L_2, of emotional experiences linked to discourse

3. Extraction, from list L_3, of emotional experiences linked to the emotional state of the child during discourse
4. Fusion of lists L_1, L_2 and L_3 into L_4 and recalculation of the coefficient associated to each emotional experience in function to:
   − Mood of the robot
   − Affect of discourse action
   − Phase and discourse act
   − Affect of the child's emotional state
   − Affect of discourse
5. Extraction of best emotional experiences from list L_4 into L_5
6. Expressions of emotions linked to chosen emotional experiences. These expressions determine the behavior of the robot.

– The "Output" Module gives the reaction expression for a system with the material characteristics it is made of: HP and motors for our robotic system.

5   Dynamic Behavior

Dynamics is set up to increase the expressiveness of the model [13]. The purpose is to give the robot a way to express itself freely, despite absence of speech or other external events. This free expression is associated to the term "dynamics". There is a link between dynamics and temperament, personality, mood, experience, etc.

Dynamics is implemented on a basic avatar to be able to make evaluations more quickly and easily. This study applies to eyebrow and lip movement - which are the same for the robot and the avatar - as well as the eye, head, and trunk movement of the avatar. As a more sophisticated version of the avatar will be released for integration on mobile devices, such as telephones and pdas, it was voluntarily given more degrees of freedom than the robot.

Dynamics is composed of three modes: waiting mode, which is the initial state of the model, listening mode and playing mode.

Waiting mode: It represents the behavior of the avatar when there is no external events. The avatar is totally free, it can move as it wants and does not pay attention to a hypothetical interlocutor.

Listening mode: It represents the behavior of the avatar in the case of interaction. For instance, it can watch a video, listen to a story or speak with an interlocutor, and then react to the various events. In this mode, it is not completely free to move because it must remain focused on the interaction.

Playing mode: It represents the behavior of the avatar when it has to express a particular emotion (requested by the interaction model following an event). Thus, the avatar loses its freedom of movement as it must express emotions as requested by the model. It continues to display emotions, one after the other, as long as there is emotion to express. This mode has priority over both of the other mode and is active as soon as an emotion has to be expressed. Since emotions can be cumulated, the display mode is automatically disactivated when there are no more requested emotions. In this case, the previous mode is reactivated.

These modes are illustrated by four dynamics parameters :

- breathing,
- eyes winking,
- gaze movement,
- face "grins".

6   Robotics Conception

This robot was partially conceived by the CRIIF for the elaboration of the skeleton and the first version of the covering. The second and third versions (see Fig. 2) were made in our laboratory. The skeleton of the head (see Fig. 2(a)), made with ABS before and epoxy resin, contains:

- 1 camera at nose level to follow the face and potentially for facial recognition. The camera used is a CMUCam 3.
- 6 motors creating the facial expression with 6 degrees of freedom. Two for the eyebrows, and four for the mouth.
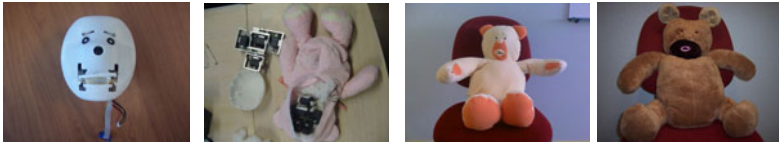
(a) Head concep- (b) Body concep- (c) EmI version 2 (d) EmI version 3
tion   tion

Fig. 2. EmI conception

The skeleton (see Fig. 2(b)) of the torso is made of aluminium and allows the robot to turn its head from left to right, as well as up and down. It also permits the same movements at the waist. There are a total of 4 motors that create these movements.

Currently, communication with the robot is done through a distant PC directly hooked up to the motors. In the short term, the PC will be placed on EmI to be able to process information while allowing for interaction.

The third version is not exactly finished, but we currently use it for preliminary evaluation and experimentation with children. Examples of expression of emotion for ArtE and EmI versions 2 and 3 are shown in Fig. 3.

Because degrees of freedom used by ArtE are not the same that EmI, dynamics of movement will be different that what we use with ArtE. From all parameters (rictus, gaze movement, eye winking ans respiration), only rictus and gaze movement will be used. Gaze movement will be convert by head movement.
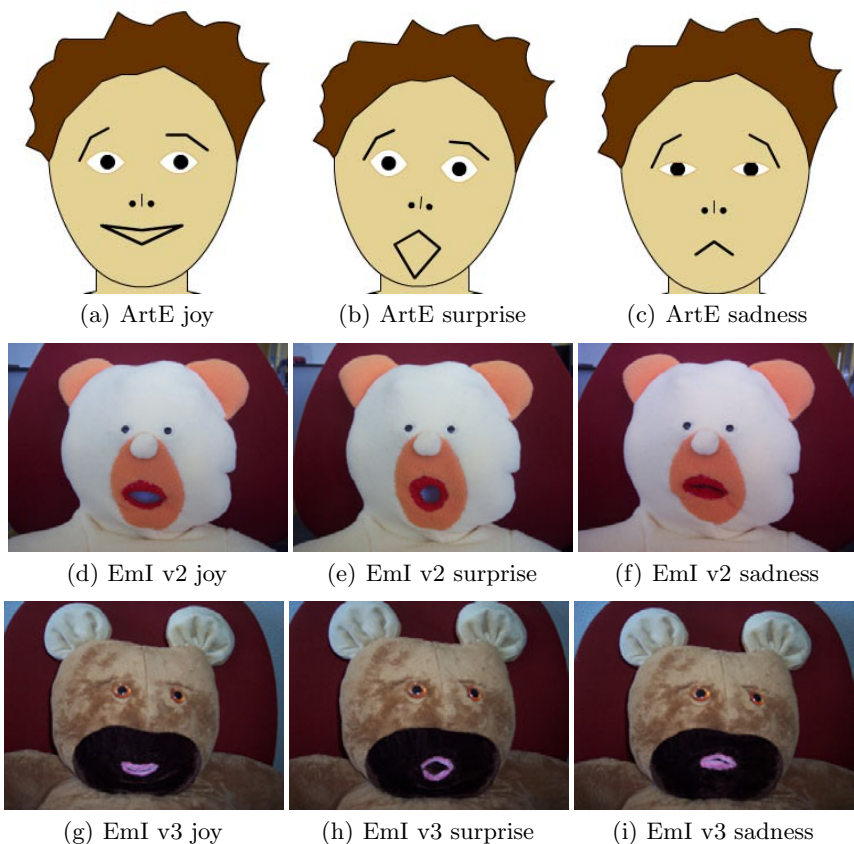
(a) ArtE joy   (b) ArtE surprise   (c) ArtE sadness

(d) EmI v2 joy   (e) EmI v2 surprise   (f) EmI v2 sadness

(g) EmI v3 joy   (h) EmI v3 surprise   (i) EmI v3 sadness

Fig. 3. Comparaison of some facial expressions

7   Conclusion

This article has presented the research we have done for the EmotiRob project. We have briefly described some of the hypotheses and models we have used for interaction between chidren and the EmI companion robot. Each previously presented module has already been evaluated separately. The results, which are not presented in this article, are very promising.

We have now began the integration of all the modules (understanding, interaction, and dynamics) for future emperimentation of interaction between EmI and children. This experimentation will allow us to validate all of the choices that have been globally made for this project.

References

1. Boyle, E.A., Anderson, A.H., Newlands, A.: The effects of visibility on dialogue and performance in a cooperative problem solving task. Language and Speech 37(1), 1–20 (1994)

2. Le-Pévédic, B., Shibata, T., Duhaut, D.: Study of the psychological interaction between a robot and disabled children (2006)
3. Villaneau, J., Antoine, J.-Y.: Deeper spoken language understanding for man-machine dialogue on broader application domains: A logical alternative to concept spotting. In: Proceedings of SRSL 2009, The 2nd Workshop on Semantic Representation of Spoken Language, Athens, Greece, pp. 50–57. Association for Computational Linguistics (March 2009)
4. Bassano, D., Labrell, F., Champaud, C., Lemétayer, F., Bonnet, P.: Le dlpf: un nouvel outil pour l'évaluation du développement du langage de production en français. Enfance 57(2), 171–208 (2005)
5. Maarouf, I.E.: Natural ontologies at work: investigating fairy tables. In: Proceedings of Corpus Linguistics conference - CL 2009, Liverpool, UK, Juillet (2009)
6. Abney, S.: Parsing by chunks. Principle-based parsing, pp. 257–278 (1991)
7. Le Tallec, M., Villaneau, J., Antoine, J.-Y., Savary, A., Syssau- Vacarella, A.: Emologus—A compositional model of emotion detection based on the propositional content of spoken utterances. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2010. LNCS, vol. 6231, pp. 361–368. Springer, Heidelberg (2010)
8. Syssau, A., Monnier, C.: Children's emotional norms for 600 French words. Behavior Research Methods 41(1), 213 (2009)
9. Petit, M., Pévédic, B.L., Duhaut, D.: Génération d'émotion pour le robot maph: média actif pour le handicap. In: Proceedings of the 17th International Conference on Francophone sur l'Interaction Homme-Machine, IHM, Toulouse, France. ACM International Conference Proceeding Series, vol. 264, pp. 271–274. ACM, New York (2005)
10. Saint-Aimé, S., Le-Pévédic, B., Duhaut, D.: Building emotions with 6 degrees of freedom. In: IEEE International Conference on Systems, Man and Cybernetics, ISIC 2007, pp. 942–947 (October 2007)
11. Saint-Aimé, S., Le Pévédic, B., Duhaut, D.: iGrace – Emotional Computational Model for EmI Companion Robot, vol. 4, pp. 51–76. Tech Education and Publishing (2009)
12. Dang, T.-H.-H., Letellier-Zarshenas, S., Duhaut, D.: Grace – generic robotic architecture to create emotions. In: Advances in Mobile Robotics: Proceedings of the Eleventh International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, pp. 174–181 (September 2008)
13. Jost, C.: Expression et dynamique des émotions. application sur un avatar virtuel. rapport de stage de master recherche, Université de Bretagne Sud, Vannes (June 2009)