Emotirob: from understanding to cognitive interaction
Amel Achour, Marc Le Tallec, Sébastien Saint-Aimé, Brigitte Le Pévédic,
Jeanne Villaneau, Jean-Yves Antoine, Dominique Duhaut

To cite this version:

HAL Id: hal-00507537
https://hal.archives-ouvertes.fr/hal-00507537
Submitted on 30 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—The ANR project EmotiRob aims at conceiving and carrying out a soft toy robot which can interact emotionally and cognitively with handicapped and fragile children. However the project MAPH (Active Media For the Handicap) which is an extension of EmotiRob extends the cognitive abilities of the robot so as to implement linguistic interaction with the child. This article presents our work for both projects: speech understanding, emotional interaction, cognitive interaction.

I. INTRODUCTION

A new stake for robotics is the domain of companion robots which execute complex tasks and offer an behavior enrichment through their interaction with human beings.

The objective of this project is to design an autonomous and “reactive” stuffed robot which might provide some comfort to vulnerable children (for example, children undergoing long-term hospitalization). Previous experiments have already shown the contribution of robot companions in this type of situation. Compared to these studies, one of the specificities of our robot is to be able to react to the behavior of the child by simulating emotions through body movements, facial expressions and small simple sounds.

This project, centered on detection and simulation of emotions, lies at the interface of robotics and human-machine communication and vision. It particularly raises several interesting scientific problems:

- “understanding” the language of a child who talks to his stuffed animal (in this case, the robot) and detecting their emotional state,
- equipping the robot with a pertinent capacity of emotional reactions, and making these emotions intelligible through the robot’s expression,
- evaluating the contribution of this interaction: how does one measure the comfort of a child, etc.

II. UNDERSTANDING SYSTEM

In the EmotiRob project, the understanding system has to handle a child’s language and detect his emotional state. To do this, we have decided to use an older understanding system, LOGUS, and to adapt it.

A. Logus

The LOGUS system [1] proceeds in three main stages, as indicated in Figure 1, to transform the statement of a logical formula. These three stages are the lemmatisation, the segmentation, and the creation of the semantic links.
domains, with any person. We know that to represent all the

In the Emotirob project, we have to create the new

The first operation was to create a new lexicon with the

The module of understanding is based on the same
architecture as the LOGUS system, as well as three stages.

At first, the idea is not to create one ontology, but much
ontology in domains very known by children, and to connect
each to others (for example animals, school, family...). It
will be necessary to verify, notably in the corpus, how the
children approach the various domains to choose what the
indispensable domains for our knowledge are.

For the moment, as we have many semantic labels, the
first work consisted in classifying all our objects in
subsections to create major categories, for example, animals,
people, places, food... These categories are then divided up
into sub-categories. For example, we classified animals into
three categories: the familiar animals (that can live in the
home), the close animals (that the children are used to
seeing) and the distant animals. These successive divisions
allow us first to decrease the number of labels to be treated
and to look for the properties common to the objects of the
same category.

This works is at the beginning and we have to continue to
create ontology. After we will have to search how detecting
emotions of children.

III. EMOTION SYNTHESIS

Emotion synthesis is an artificial reproduction of the way
that human beings express their emotions in a social human
context. Being able to express emotions is a way for a robot
to communicate with a human. Without emotional
expressions, users may interpret the interaction as a lack of
interest from the robot and will then stop communication.

Robots (like Kismet, Paro, Necoro, Aibo, Asimo, etc.)
capable of social interaction are then characterized by their ability to have behavior close to that of humans.

Emotion is an expression of one’s internal state. Joy and fear are different emotional states and can be characterized with physiological and psychological criteria. Through emotion synthesis, those criteria are imitated to bring about a believable artificial emotion. This artificial emotion is then used to reinforce communication: verbal or non verbal.

A. Facial Animation for EmotiRob

We have presented research on the question “how can an emotion be simply drawn?” In this evaluation, based on a specific set of faces having different degrees of freedom [13], it can be noticed that the one with the highest score is not the most complex one, as in Fig. 2.

After defining the face for our robot it will now be necessary to give it dynamics. Knowing that we are primarily working with emotional experiences [18], the first task will be to define the emotions felt for each of these experiences. These emotions will be classified and organized into a hierarchy according to several existing models [14] [15] [16] [17] [18]. This categorization allows us to choose the best method to go from one emotion to another.

**Fig 2. Face and examples of expressions for the robot**

**IV. MECHANICS OF THE ROBOT**

The mechanics of EmotiRob can be separated into three modules. The first module is the head. It contains all of the motors in order to set the eyebrows, jaws and corners of the mouth into motion. The second module is the neck, it contains two motors. The third module is the pelvis which also contains two motors

A. Activation of the eyebrows

The eyebrows are operated independently by two servo-motors type BMS 303 (BlueBird). The actuator results in a rod-crank system guided by a groove in the structure. The groove is straight for the moment in Fig. 3, but it might also be curved to enhance the expression of the robot.

**Fig 3. Head of EmotiRob**

B. Activation of the jaws

The upper and lower jaws are independent from each other. Each is driven by a servo-motor type BMS 303 (BlueBird). The range of motion of the jaw should not be too large.

C. Activation of the lips

The corners of the lips are also independent from each other to increase the possibilities of facial expressions. Each is operated by a servo-type BMS 380 (BlueBird), which pull a cable on which a two-spring traction (upper and lower) is fixed. These springs allow for the various lengths of the "lips" while playing the role of tendon from the mouth. The cables pass through rings along the upper jaw. This will transform the movement of rotation of the actuator movement in translation along the jaw. The BMS 380 servo-motors are sized to be able to stretch the springs.

D. Neck and pelvis mechanics

Both modules are treated the same for reasons of ease of implementation and since their features are the same. This module must be disengaged: the force on the rotation should not damage the motors. With force rotation, the compression spring shrinks as a result of a force that is applied along the axis of rotation. This contraction disassociates (disengagement) both toothed surfaces and thus the rotation takes place freely without damaging the mechanical actuator. The principle is the same for both axes. Once there is no more force on the rotation, the compression spring applies an effort (clutch), which goes put in touch toothed surfaces.
The Cognitive interaction module corresponds to a part of the MAPH project (Média Actif Pour le Handicap) which is related to the EmotiRob project. It aims at extending the reaction capacities of the robot so that it could maintain a "conversation" in a natural language with the child. The purpose of this work, indeed, is to build a linguistic and cognitive interaction module between the child and the robot by generating new sentences to carry out a "conversation" with the child. The understanding module presented below supplies inputs for the cognitive interaction. The sentences generated by the robot depend on the emotional states of both robot and child.

Carrying on a "natural" conversation with a machine on a non-constrained subject seems to be very difficult and even impossible [23] as we cannot model world knowledge right now. By restricting the field of the conversation and choosing a well-defined subject, the use of natural language in order to carry a "dialogue" becomes feasible but still very difficult. Some existing systems of human-machine dialogue that we can quote here are COALA1 which is a system for documentary assistance in a town media library, CMU Communicator2 which is realized by Carnegie Mellon University, and serves as a tourist guide. Problems that come up against the conception and the realization of human-machine dialogue systems are essentially oral language recognition and understanding [21], real time conversation constraint and finally, speaking with the machine must not require learning from the user [22].

Regarding our work, we are implementing a generic human-machine dialogue system dedicated to young children. We are dealing with vocabulary covering the child's entire surroundings. This is quite problematic as we have a non-restricted conversation domain. However, limiting the users of our system to young children makes the vocabulary we are interested in quite restricted. Under this condition, producing a dialogue between the child and the robot is conceivable.

In this work, we used a corpus elaborated by means of the DLPF tool, realized by D. Bassano, F. Labrell, C. Champaud, F. Lemetayer and P. Bonnet [19] and used to estimate the development of production language of French children whose age is between two and five. The corpus counts a little less than 1500 words including nouns, verbs, adjectives, adverbs, onomatopoeias and common expressions, articles and pronouns... among which we find all the common vocabulary that could be said by a 4-year-old child. The problem which arises is then how to model the conceptual world of a very young child. By studying the semantic relations which could exist between the different words of the corpus, we established a classification of the words or rather taxonomy according to not only objective but also emotional properties. Moreover, we tried to be in accordance with a child's way of thinking and perception of things. That is why we tried to validate our research with children and to verify if we have satisfied these constraints.

A. Taxonomy creation and validation with children

In order to calculate the coefficients of the semantic link between pairs of words, we start with classifying the corpus words in various classes and sub-classes according to their meanings. The figure below shows a small outline of the taxonomy that we obtained. The detail of the taxonomy is visible on the website of the EmotiRob project [24].

![An outline of the taxonomy of children words](image)

To validate the words taxonomy that we created, we thought of making a questionnaire and had children between the ages of 5 and 7 fill it out. According to Piaget [20] the children of the lower ages tend to fantasize and to say anything when they do not know the answer of the question or even when they are not sure of it. It is what Piaget called the "n’importequisme"3 phenomenon.

To implement the various questions and parts of the questionnaire, we used the test method, which is often

---

1 http://www-ic2.univ-lemans.fr/~lemeunie/these/node35.html
2 http://www.speech.cs.cmu.edu/Communicator/
3 "whateverism"
employed in the study of childhood beliefs. This method requires two essential conditions: the first one is that questions should be the same for all the subjects and asked in the same circumstances. The second condition requires that all the answers must be reported to the same evaluation scale. This method is more efficient if someone varies the questions and makes counter-suggestions. Thus, in our questionnaire we opted for various types of questions such as questions with multiple choices, tables to fill out, as well as searching for adequate solutions among a set of possible ones.

This questionnaire essentially concerns the animated beings that surround the children: human or animal, and with whom they can have more or less emotional links. The first part of the questionnaire was dedicated to human beings. At first, we asked the children to distinguish the characters we find around us from those who exist only in tales. And secondly, we asked them to reveal the different primary feelings which these characters evoke for them.

The second part was dedicated to animals and aimed at verifying their belonging to the different categories established in the taxonomy. The questionnaire was filled in by first-grade pupils from "Sainte Marie" elementary school in Lorient in France twice. After the counting of the answers, we noticed that, in the majority of the cases, the results are in accordance with the taxonomy and answer, indeed, our expectations. Nevertheless, certain results surprised us and made us modify the taxonomy. For example, according to 92% of the questioned children, "king" and "prince" characters do not exist in the real world and are only in tales and stories we tell them. "Father Christmas" belongs to the set of imaginary characters for only the half of the children, the other half consider him as a real person. 80% of children are "happy" when they see a "magician" and consider him as a "kind person". To many children, almost half, an "ostrich" and a "penguin" are not "magician" and consider him as a "kind person". To many children, nearly 70% of them think that a "whale" is a fish.

**B. Calculation of semantic rapprochement coefficients**

Using the taxonomy, we defined and measured semantic rapprochement coefficient which exist between the corpus words. In what follows, we describe these coefficients, as well as the method used for calculation.

**Rapprochement coefficient between two words having the same type:** The set of common nouns was divided into three big classes, the class of animate, inanimate and abstracts. Each of these classes was afterward divided in several sub-categories, and so on. Here is the formula used to calculate the distance $Rapp(N1, N2)$ between two common nouns $N1$ and $N2$:

$$Rapp(N1, N2) = \frac{C1*R1(N1, N2) + C2*R2(N1, N2)}{C1 + C2}$$

As can be noticed, the rapprochement coefficient between two common nouns is a weighted average between two coefficients, the first of which, $R1(N1, N2)$, calculates the rapprochement between both words in the taxonomy, whereas the second evaluates their rapprochement regarding to their common properties number. We distinguished two types of properties: affective properties and objective ones. Each property was balanced with a weight measuring its importance in defining a certain set of words. $R2$ is then the weighted average of an affective rapprochement $Raff(N1, N2)$ balanced by an affective coefficient $Qa$, and an objective rapprochement $Robj(N1, N2)$ balanced by an objective coefficient $Qo$.

$$Raff(N1, N2) = \frac{nbr\_prop\_aff\_com(N1, N2)}{max(nbr\_prop\_aff(N1), nbr\_prop\_aff(N2))}$$

$$Robj(N1, N2) = \frac{nbr\_prop\_obj\_com(N1, N2)}{max(nbr\_prop\_obj(N1), nbr\_prop\_obj(N2))}$$

$$R2(N1, N2) = \frac{Qa*Raff(N1, N2) + Qo*Robj(N1, N2)}{Qa + Qo}$$

The rapprochement coefficients we obtained depend on the $Qa$ and $Qo$ that we chose. For instance, "ladybug" and "louse" will be semantically close if $Qo$ is bigger than $Qa$. Otherwise, they will be distant.

The semantic rapprochement between two verbs or two adjectives is calculated in the same way. However, for verbs, we take their respective types (intransitive, transitive or double transitive) and the taxonomic rapprochement between their subjects and respective complements into account. As for adjectives, we take the types of the subjects that could be applied to them into account.

**Rapprochement between verb and noun, adjective and noun:** We defined the rapprochement coefficient between noun and verb to measure the applicability of some verbs to a particular noun. For example, we can say that the verb "to bore" applies perfectly to an animate subject whereas if the same verb can be applied to a "chair", that is possible only in an artistic or a funny context. Also for adjectives, the rapprochement coefficient between noun and adjective measures rather the use of an adjective with a certain name in a particular speech context.

**C. Sentence generation**

The sentence generation model that we adopted works with simple input sentences such as (subject, adjective), (subject, intransitive verb), (subject, transitive verb, complement) or (subject, verb double transitive, complement1, complement2). At the present time, we have limited our choice to one type of sentence which is the affirmative sentences. Later, we intend to work on the acts of language and to introduce the interrogative, imperative sentences, etc. The generation module takes several parameters into account which describe the emotional state of the child, as well as the humor of the robot. First of all, it analyzes the input sentence so as to find the context that the speech is about. Secondly, it looks for words which represent a certain value of the semantic rapprochement coefficients with the words composing the input sentence. Several cases are possible, according to the speech context.
(realistic, funny, artistic, etc.) and according to the emotional states of both the child and the robot. The robot will then be able to answer the child either by keeping the same subject of discussion or a similar one, or by approaching another completely different subject.

Regarding this work, we planned for the possibility of corpus enrichment and the addition of new knowledge. So that, the system must be capable of semantically connecting the new words added to those which already exist in its knowledge database and to draw up their lists of properties.

VI. CONCLUSION

Having successfully completed our studies on the face of the robot, his dynamics and the expression of his emotions, all of the rest still needs to be modeled. We, therefore, propose a model of emotion which is not based on the unique representation of simple emotional experiences.

Most of the work now is not only the association between these emotional experiences and keywords in the discourse but more importantly develop the process of generating emotions and behavior of the robot.

Our future work will focus on the development, as well as the evolution and the improvement of the mechanical robot to increase its functionality.

As for the cognitive interaction, we intend to make the system able to enrich its own vocabulary, and to define learning frameworks; according to the answers the system is going to produce, in order to adapt it to our expectations and to those of the children.

ACKNOWLEDGMENT

EmotiRob is currently financed by the regional council of Martinique, for the development of the emotional synthesis part, the regional council of Bretagne for the language comprehension part, and the ANR for the construction of the robot.

First of all, we would like to thank the regional council of Martinique, as well as the ANR for their collaboration.

The authors would like to thank Claire PALLARD, pediatrician, for her contribution for the realization of an evaluation grid.

The authors would like to thank members of the center of Kerpape (Laurence TRIVIERE, Marcellle SAMSON, Sylvie LEMONNIER and Annie BERTRAND) for their cooperation with our experiment.

The authors would also like to thank members of the IEA “Le Bondon” (Florence CHARPENTIER and Olivier THUAU) for their cooperation with our experiment.

The authors would like to thank Dominique Bassano for her corpus which allowed us to have such a complete tool on the vocabulary spoken by young children.

REFERENCES