A Wizard-of-Oz Game for Collecting Emotional Audio Data in a Children-Robot Interaction

Agnes Delaborde
LIMSI-CNRS
BP133 91403 Orsay cedex
+33169858062
agnes.delaborde@limsi.fr

Marie Tahon
LIMSI-CNRS
BP133 91403 Orsay cedex
+33169858062
marie.tahon@limsi.fr

Claude Barras
LIMSI-CNRS
Université Paris Sud
BP133 91403 Orsay cedex
+33169858061
claude.barras@limsi.fr

Laurence Devillers
LIMSI-CNRS
Université Paris Sud
BP133 91403 Orsay cedex
+33169858062
devil@limsi.fr

ABSTRACT

In this paper, we focus on the recording protocol for gathering emotional audio data during interactions between the Nao robot and children. The robot is operated by a Wizard-of-Oz, according to strategies meant to elicit vocal expressions of emotions in children. These recordings will provide data to develop a real-time emotion detection module for the robot, and will be a starting point for a study on emotional models of exchanges in the interaction with a robot. This work is carried out in the context of the French ROMEO project meant to design a robot which will be able to interact with people (children, elder people), taking into account the behavior of the person. Two kinds of application are studied in this project: the robot acts as a game companion or as an assistant to disabled persons.

Categories and Subject Descriptors

H.2.8 Database Applications. H.5.2 User Interfaces.

General Terms

Experimentation, Human Factors, Theory

Keywords

Social robots – Real-time Emotion detection – Affective interaction – Wizard of Oz – Human-Robot Interaction

1. INTRODUCTION

In order to design affective interactive systems [1], experimental grounding is required to study expressions of emotion during interaction. The creation of an emotional corpus based on real-life or realistic data is extremely difficult to build, due to several constraints (cost, privacy) [2].

AFFINE '09, November 6, 2009 Boston, MA

This study is carried out in the context of the French ROMEO project[1], meant to design a robot which will be able to assist disabled persons in their everyday activities (carrying an object, cleaning the house, reading mails…), and to entertain and take care of children (playing games, calming down their arguments). We focus in this article on the recording protocol for gathering emotional audio data during interactions between the Nao robot (http://www.aldebaran-robotics.com) and children playing games. The robot is operated by a Wizard-of-Oz, according to strategies meant to elicit vocal expressions of emotions in children. These recordings will provide data to develop a real-time emotion detection module for the robot, and will be a starting point for a study on emotional models of exchanges in the interaction with a robot. Three types of game were designed: a Trivial Pursuit-like game, a game of songs and an emotion detection game. The challenge of this data collection is to obtain naturalistic children's behaviors presenting many emotional reactions. Several strategies have been used to try and induce spontaneous emotional reactions during these games. We also collected "prototypical" emotion expressions through the game of emotion detection allowing a comparison with spontaneous expressions. This paper mainly describes the recording protocol (Section 2) and an analysis of the collected data (section 3). Section 4 concludes this first experiment.

2. RECORDING PROTOCOL

2.1 The Participants

The players are the robot and two children (we plan to record children from 7 to 12-13 years old). The game is led by the game master who collaborates in the experiment and stimulates the interaction with the robot. The robot is remotely operated by a WoZ, according to a predefined scenario, which is also known by the game master. Another experimenter is dealing with the recordings.

Children are recorded with high quality lapel-microphones (AKG PT40 Pro Flexx), and filmed with one camera (see Figure 1). The video recording is intended for supporting the perceptive annotation of the audio data, in case of doubt.

[1] This work is financed by national funds FUI6 under the French ROMEO project labeled by CAP DIGITAL competitive center (Paris Region).

2.2 The Game

The experiment is carried out in three phases: first of all, a Trivial Pursuit-like game, during which each player asks by turns a question that is written on a card, and the two other players have to find the answer. The robot can also 'read' the question (pre-recorded) to the children. Then comes a game that consists in discovering the song that is hummed or sung by one of the children (data used for another application in ROMEO). The final stage is a game during which the robot acts as an emotion detection tool under development : the children have to try and express the emotion that is asked by the game master, until the robot detects it properly.

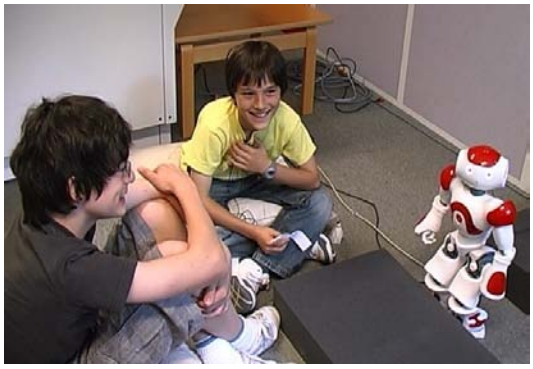

Figure 1. Two children with the robot showing expression of amusement.

2.3 Scenario Followed by the WoZ

Three emotional states are expected from the children: dissatisfaction (irritation, anger), satisfaction (joy, contentment and amusement), deception (stress, sadness). Each emotional reaction is expected to be triggered by the robot's actions, classified according to its mental states' representation [3, 4] (cf. table 1).

Table 1. Description of a possible action of the robot.

Mental state: Certainty			
Physical action	**Emotion**	**Utterance**	**Circumstances**
Raises energetically both arms	Excited, enthusiastic	"I know the answer!"	The robot rushes to give a wrong answer to an obvious question

In order to obtain various and intense enough vocal expressions of emotions, we need the robot to make fun of the human or undergo dysfunctions, such as inappropriately shutting down, talking nonsense, etc. Here are the types of strategies we will use, and the expressions of emotion we expect:

- Interaction prompting → *neutral, soft amusement*
  - Directly asks for attention ("Do you want to play?", "Do you need some hints?")
  - Motivates (congratulates the child)
- Poor understanding of its environment → *amusement from soft to high, irony, satisfaction*
  - Disregards the rules (gives the answer instead of a clue, reads the question plus the answer)
  - Is unconscious of its failures ("I'm the best !" when he plainly made a mistake)
  - Crashes (unexpected shutdown in the course of the game)
  - The robot's emotion detection tool always detects the emotion the child is supposed to express (even though the expression of emotion clearly does not fit the expected one)
- Persisting in making mistakes → *irritation (from soft to average), soft stress*
  - Asks the same questions without taking the answers into account
  - The child never manages to activate the robot's emotion detection tool (even though the expression of emotion is correct)
  - The robot does not want to reboot, in spite of the children's calls

We also expect spontaneous expressions of emotions to occur through the interaction between the children in the course of the game, although we cannot plan them.

3. EXPERIMENTATION

3.1 Emotional Data Collected

A first test recording has been realized with two male children (see Figure 1). This test allows us to observe the way the children react to the setting of the experiment, and to put to the test some of our predictions about the reactions of the children. Here are some of the most relevant situations, and the expressions of emotion they triggered in the children:

- The robot contests the rules given by the game master; children's reactions: *smiles, soft laughs*
- The robot asks for explanations, and does not seem to understand them (repeats itself like a machine that crashes); children's reactions: *patience, soft amusement, soft irritation*
- The robot reads the question on the card… and the answer as well; children's reactions: *soft laughs, patience, soft irritation*
- The robot wants to give a clue, but gives away the answer; children's reactions: *laughs*
- A child expresses correctly an acted Anger, but the robot never returns "Anger"; children's reactions: *amusement, then deception, then irritation*

3.2 Results Analysis and Further Improvements

This test allows us to question the whole setting of the game, and the choices we made concerning the elicitation strategies.

The children seem to be willing to patiently explain to the robot its mistakes, which prevents the expression of irritation that we initially expected. However, this might be due to the sole fact that

they know that the robot is under development, and thus are less particular about its technical failures.

The third game (in which children are asked to express determined emotions and the robot returns the emotion it supposedly detected) mainly triggers over-expressed acted emotions, but it also allows us to get emotions that are directly driven by the strategies used by the WoZ, which can be considered to be natural.

We find it essential to exaggerate each reaction of the robot: e.g. annoying the children by insisting, repeating itself, or alarming them (when the robot goes off and does not want to switch itself on anymore). As the robot is expected eventually to be able to detect an argument, it seems to be compulsory to get good samples of these emotions. Nonetheless, we only obtain very mild irritation or stress from the children in our test.

In order to keep the interaction going, the robot has to react rapidly enough, and the game master helps to make the interactions more dynamic. The setting of the experiment does not only concern a pure Human-Robot interaction, since a human plays the role of the intermediary. The impact of this factor will have to be subsequently examined.

We are currently working on a final improvement of the protocol, in which we will propose a wider panel of actions for the robot, and a closer feedback to the children's reactions.

4. OUTLOOK

Next steps will be to collect sufficient children-robot interactions for training emotional models (50 children are expected to participate in this experiment) and also to annotate the collected data with emotional labels and contextual information [5, 6]. This corpus will present acted, induced and spontaneous emotional audio data for the same subject, allowing potential comparative studies.

We also intend to analyze the impact of the emotional state over the children identification through their voice. Speaker identification is usually performed using statistics of short-term spectral features, which also contain information on the emotional expression.

However, the data we will collect only concern one aspect of the robot's role, i.e. entertainment. In a context of assistance, the robot will have to be more responsive to stress in the voice. We thus plan to collect another corpus within the framework of assistance to the disabled. These emotional audio corpora will allow us to train a real-time emotion detection module in the robot on various contexts and on various types of people.

These data will also provide a first basis for a study on a possible modeling of emotional exchanges in Human-Robot Interaction. The present protocol allows to test our elicitation strategies, and to find a basic model of the succession of emotions in a Human Robot Interaction. We will test in the future the hypothesis that if the human reacts as expected to the strategy of elicitation we put into practice through the robot, it may mean that we managed to get them emotionally engaged in the interaction. Since our protocol is specifically intended for the gathering of emotional audio data, it does not allow for specific measures of engagement as it is, be it a spatial measure [7] or a measure on joint attention and behavior of the robot [8]. Further experiments will focus on the analysis of the way one can increase the desire of the human to maintain the interaction with the robot, by controlling the affective link.

5. REFERENCES

[1] Picard, R. Affective Computing, MIT Press, 1997.

[2] Douglas-Cowie, E., Campbell, N., Cowie, R. and Roach, P. : *Emotional speech; Towards a new generation of databases.* In Speech Communication 40: 33-60. 2003.

[3] Zara, A. , Maffiolo, V., Martin J-C., Devillers L., *Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics*, ACII 2007.

[4] Baron-Cohen, Simon, Helen Tager-Flusberg, and Donald J. Cohen, eds. *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford: Oxford University Press, 2000.

[5] Devillers, L., Abrilian, S. and Martin, J.C. *Representing real life emotions in audiovisual data with non basic emotional patterns and context features.* Presented at 1st Int. Conf. Affective Computing and Intelligent Interaction (ACII'2005), Beijing, China, 2005.

[6] Devillers, L., Martin, JC. *Coding Emotional Events in Audiovisual Corpora*, LREC 2008.

[7] Michalowski, M. P., Sabanovic, S., and Simmons, R.. *A spatial model of engagement for a social robot.* In Proceedings of the 9th IEEE International Workshop on Advanced Motion Control, Istanbul, Turkey, 2006.

[8] Sidner, C. L., Kidd, C. D., Lee, C. H., and Lesh, N. B. *Where to look: A study of human-robot engagement*. In Proceedings of the 9th international conference on Intelligent user interfaces, pages 78–84, Funchal, Madeira, Portugal, 2004.

[9] Ochs, M., Pelachaud, C., Sadek, D. *An Empathic Virtual Dialog Agent to Improve Human-Machine Interaction***.** Autonomous Agent and Multi-Agent Systems (AAMAS), Estoril, Portugal, 2008.

[10] Nomura, T., Suzuki, T., Kanda, T., and Kato, K. *Measurement of Anxiety toward Robots.* In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, University of Hertfordshire, Hatfield, United Kingdom, 6-8 September 2006.

[11] Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti & Peter Robinson. *Empathizing with Robots: Fellow Feeling along the Anthropomorphic Spectrum.* Affective Computing & Intelligent Interaction conference, Amsterdam, septembre 2009.