Corpus NAO-Children : Affective Links in a Child-Robot Interaction

Agnes Delaborde, Marie Tahon, Claude Barras, Laurence Devillers

LIMSI-CNRS
BP 133, 91 403 Orsay cedex, France
(agnes.delaborde|marie.tahon|claude.barras|laurence.devillers)@limsi.fr

Abstract

1. Introduction

The NAO-Children corpus is being collected in the course of the Project ROMEO (Cap digital French national project founded by FUI6, http://www.projetromeo.com). The aim of the project is to design a robotic companion which can play different roles : a robot assistant (1m40 – 4.6') for dependant people (visually-impaired, elderly people) and a game companion with children. Our first experiments have been done with the prototype Nao (58cm – 23") remotely operated with a Woz (Wizard of Oz). We focused on the role of the robot as a game companion.

In this role, the robot has to be able to supervise a game, while also being sensitive to the emotions of the children. It should for example be able to detect through the expressed emotions if a child is sad, if there's a tension between the children. It also has to behave in a way such as to entertain the children, and maintain their desire to play with it.

The NAO-Corpus will allow to train models on real-time emotion detection in children's voice, in a Human-Robot Interaction context. This will also allow studies on speaker identification

AIM : emotion detection in Human-Robot Interaction, speaker identification robust to emotional speech, and emotional interaction models of children at play with a robot.

Emotional system : Can emotion detection add context to an emotional system ?

NAO-Children Corpus : description of the corpus, and its annotations

Affective Links : study of affective links on this corpus : H-H and H-R interaction.

2. Emotional System

The degree of human-likeness a robot should reach is bounded by perceptive characteristics among humans: although users expect a human-like robot to bear psychological and intellectual resemblances with a human (Hegel et al., 2008), the robot must remain credible enough so as not to appear unfamiliar (Mori, 1970; Ho et al., 2008; Nomura et al., 2006). If a robot speaks your language, waves at you, looks at you, in one word interacts with human means, you could expect it to be able to have some other human communicative features, such as being able to understand your emotions and react to them (Riek et al., 2009; Nomura et al., 2006; Thomaz et al., 2005).

An emotion-sensitive robot acting in real-life conditions would need to adapt to new situations, new persons, and be able to understand the human's personality, mental state and social relations. So far, models used for emotional systems are mostly derived from storytelling tasks, which means that these models take into account an almost full context of interaction : be it the non-human entity's past and memory (Ochs et al., 2009; Bickmore et al., 2009), its relation to the other humans or non-human entities around it (Rousseau and Hayes-Roth, 1998; Kaplan and Hafner, 2004; Michalowski et al., 2006; Sidner et al., 2004), or a modeling of the events happening in the agent's surroundings (Ochs et al., 2009), etc.

However, emotions in real-life conditions are complex, and factors responsible for the emergence of an emotional manifestation are intricate (Scherer, 2003). Detecting automatically the emotion out of the human voice is in itself a real challenge : without any clues about the context of emergence, the events or emotional dispositions of the speaker, the detection only relies on prosodic features [ref ?]. What can be infered of the human's personality and his or her relation to the robot from emotion detection in a human-robot interaction ? Can emotion detection add context to an emotional system ?

An audio corpus annotated with interaction and emotional information will provide a basis for a survey of the audio characteristics that can be linked to an emotional profile of the speaker.

3. NAO-Children Corpus

Designing affective interactive systems needs to rely on an experimental grounding (Picard, 1997). However, cost, or privacy, are disuassive in the creation of an emotional corpus based on real-life or realistic data (Douglas-Cowie et al., 2003).

In the NAO-Children corpus (Delaborde et al., 2009), two children by session are recorded as they play with the robot. A game master supervises the game, gives the question cards, and encourages the children to interact with the robot. In order to reinforce the emotional reactions of the children, only friends or sisters and brothers are recorded.

The robot in this context is a player, and also tries to answer the questions. In order to trigger emotional reactions in the children, it acts in an unconform way from times to times.

3.1. Corpus Characteristics

So far, ten French children (five girls, five boys), aged between eight and thirteen years, have been recorded with high-quality lapel-microphones for a total amount of about two hours of recordings.

3.1.1. Objectives of the Corpus

The corpus aims at gathering recordings of children playing in a family setting (brother and sisters, or friends). It provides acted, induced and spontaneous emotional data. These data will be a GROUND for studies on emotion detection in Human-Robot Interaction, speaker identification robust to emotional speech, and emotional interaction models of children at play with a robot.

3.1.2. Protocole

Children were offered to play games with the robot.
1) Questions-Answers game : each player reads by turn a question written on a card, and the two others try and guess the answer.
2) Songs game : each human player in turn has to hum the song which title is written on a card, until the robot recognizes the song.
3) Emotions game : each human player in turn acts out an emotion (Fear, Joy, Anger, Sadness), until the robot recognizes it.
The robot is remotely operated by a Wizard of Oz, who loads predetermined behaviors meant to have the children react. In the course of the game, the robot plays several roles. It will be an attentive game player and quietly answer or help to answer the questions. But it will also crash unexpectedly, refuse to help a child, favors one child rather than the other, mix up the rules.

3.1.3. Annotation Scheme

On each child's track, we define segment's boundaries. A segment is emotionally homogenous, i. e. the emotion is the same and of a constant intensity along the segment. Each segment will be annotated by different professional labelers.

- Emotion Label

So as to describe the complexity of the expressed emotions, three emotion labels (cf. Table 1) are used to describe each segment. Every combinings are possible : positive labels can be mixed with negatives ; there can be no perceived emotions at all.

- Mental State (Zara et al., 2007 ; Baron-Cohen et al., 2000)

By observing the speaker talking, his or her emotional reaction, what can we infer about his or her thoughts, desires or intentions ? e. g. *to be sure*, *to doubt*, *to agree*, etc.

- Valence

Does the speaker feel a positive or a negative sensation ? *positive*, *negative*, *ambiguous : either positive or negative*, *positive and negative*, *valence unperceivable*

- Trigger Event

What kind of event triggered the child's emotional reaction ? If the trigger comes from the other child, what type of communication act ? e. g. *encourages*, *laughs at*, *laughs with*, *explains*.

Emotion's Category	Annotation value
POSITIVE	Joy Amusement Satisfaction Empathy Compassion Motherese Positive
NEGATIVE	Anger Irritation Scorn Negative Boredom
SADNESS	Sadness Disappointment
FEAR	Fear Anxiety Stress Embarrassment
NEUTRAL - AMBIGUOUS	Neutral Surprise Interest Irony

Table 1: Emotion labels

If the trigger comes from the robot, what type of elicitation strategy ? e. g. *asks for attention*, *encourages*, *inappropriately goes off*, *never recognizes the child emotion/song*.
If the trigger comes from the game master, what type of communication act ? e. g. *child control - security*, *explains the rules*, *the game master reinforces an elicitation strategy that failed*.

- Intensity

The strength of the expressed emotion. Scale of 1 to 5, from *very weak* to *very strong*.

- Activation

How many phonatory means are involved in the expression of the emotion (sound level, trembling voice, overarticulation, etc.) ? Scale of 1 to 5, from *very few* to *a lot*.

- Control

Does the speaker control, contain his or her emotional reaction ? Scale of 1 to 5, from *not at all* to *completely*.

- Spontaneity

Does the speaker reacts spontaneously, or have they been asked to act out an emotion ? *spontaneous* or *acted*.

4. Affective Links

The NAO-Children corpus gathers annotated emotional data, which allow us to test the affective interaction strategies applied through the robot. The result of these strategies will allow to define and validate some emotional profile among children in interaction with a humanoid robot,

and the usual emotional reactions linked to one profile or another. The interest of this data collection is twofold : we observe children interacting with each other, and with the robot.

In a real gaming interaction, the corpus could provide a basis for the determination of the robot's role and associated behaviors according to the average expressed emotions. For example, when the robot makes mistakes, young girls tend to mother the robot and explain it patiently (involvment in the interaction). On the other hand, teenage boys seem more inclined to condescend to it, and make use of irony (disengagement). Both girls and boys sometimes simply laugh at the robot's obvious mistakes. Can we draw a global rule to predict what will please a type of child, and what will not ? To what extent will a defect be considered as funny and attractive ?

The affective links between the robot and the child can be balanced by the social relationship between the children. Would an older brother be more willing to take control of the game and assert his dominance on his brother and on the robot ? In what way should the robot act in consequence ?

Studies on social interaction between human and non-human entities, and its influence on emotion expressions (Ochs et al., 2009, Bickmore and Cassell, 2005, Isbister, 2006), could allow us to determine some tendency of the speaker's emotional dispositions. By allowing an adaptive and dynamic processing of the affective links between the human and the robot, emotion detection can bring valuable indications to maintain the interaction.

5. Conclusion

more children will be recorded

towards an Interactive Stories game (user takes part in the construction of the story ; takes into account the emotional reaction and the profile of the user in the course of the game)

6. References

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

Timothy Bickmore and Justine Cassell. 2005. Social Dialogue with Embodied Conversational Agents. *Advances in Natural Multimodal Dialogue Systems*, 30:25–54.

Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2009. Engagement vs. Deceit: Virtual Humans with Human Autobiographie. In *Proceedings of Intelligent Virtual Agents*, Amsterdam, The Netherlands.

Agnes Delaborde, Marie Tahon, Claude Barras, and Laurence Devillers. 2009. A Wizard-of-Oz Game for Collecting Emotional Audio Data in a Children-Robot Interaction. In *Proceedings of the International Workshop on Affective-aware Virtual Agents and Social Robots, ICMI-MLMI*, Boston, USA.

Ellen Douglas-Cowie, Nick Campbell, Roddy Cowie, and Peter Roach. 2003. Emotional speech; Towards a new generation of databases. *Speech Communication*, 40:33–60.

Frank Hegel, Soren Krach, Tilo Kircher, Britta Wrede, and Gerhard Sagerer. 2008. Understanding Social Robots: A User Study on Anthropomorphism. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man08)*, pages 574–579, Munich, Germany.

Chin-Chang Ho, Karl F. MacDorman, and Z. A. D. Dwi Pramono. 2008. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In *Proceedings of the Third IEEE International Conference on Human-Robot Interaction*, pages 169–176, Amsterdam, The Netherlands. ACM.

Katherine Isbister. 2006. *Better Game Characters by Design: A Psychological Approach*. Morgan Kaufmann Publishers Inc.

Frederic Kaplan and Verena V. Hafner. 2004. The Challenges of Joint Attention. In *Proceedings of the Fourth International Workshop on Epigenetic Robotics*, Genoa, Italy.

Marek P. Michalowski, Selma Sabanovic, and Reid Simmons. 2006. A Spatial Model of Engagement for a Social Robot. In *Proceedings of the 9th IEEE International Workshop on Advanced Motion Control*.

Masahiro Mori. 1970. Bukimi no tani. The Uncanny Valley. *Energy*, 7(4):33–35.

Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of Anxiety toward Robots. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man06)*, pages 372–377, University of Hertfordshire, Hatfield, United Kingdom.

Magalie Ochs, Nicolas Sabouret, and Vincent Corruble. 2009. Simulation of the Dynamics of Non-Player Characters Emotions and Social Relations in Games. In *Transactions on Computational Intelligence and AI in Games*.

Rosalind W. Picard. 1997. *Affective Computing*. MIT Press.

Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. Empathizing with Robots: Fellow Feeling along the Anthropomorphic Spectrum. In *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction*, Amsterdam, The Netherlands.

Daniel Rousseau and Barbara Hayes-Roth. 1998. A social-psychological model for synthetic actors. In *Proceedings of the International Conference on Autonomous Agents*, pages 165–172, Minneapolis, MN, USA. ACM.

Klaus Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40:227–256.

Candace L. Sidner, Cory D. Kidd, Christopher Lee, and Neal Lesh. 2004. Where to Look: A Study of Human-Robot Engagement. In *Proceedings of the 9th international conference on Intelligent user interfaces*, Funchal, Madeira, Portugal.

Andrea Lockerd Thomaz, Matt Berlin, and Cynthia Breazeal. 2005. An Embodied Computational Model of Social Referencing. In *Proceedings of the 14th IEEE In-*

ternational Symposium on Robot and Human Interactive Communication (Ro-Man05), Nashville, TN, USA.

Aurelie Zara, Valerie Maffiolo, Jean-Claude Martin, and Laurence Devillers. 2007. Collection and Annotation of a Corpus of Human-Human Multimodal Interactions: Emotion and Others Anthropomorphic Characteristics. In *Proceedings of the second International Conference on Affective Computing and Intelligent Interaction*, Lisbon, Portugal.