Nao is doing humour in the CHIST-ERA JOKER project

Guillaume Dubuisson Duplessis, Lucile Béchade, Mohamed Sehili, Agnès Delaborde, Vincent Letard, Anne-Laure Ligozat, Paul Deléglise, Yannick Estève, Sophie Rosset, Laurence Devillers

2.1. Riddles

One of the goals of the JOKER system is to keep the social interaction entertaining by telling riddles. The riddles follow a common structure. First, the system asks a question to form the riddle. Then, the human participant suggests an answer (in general, riddles are made in a way that the answer is not expected to be found). Finally, the system provides the right answer and makes a positive or negative comment (cf. section 2.3).

The riddle database of the system is structured into four categories: (i) social humour (socially acceptable riddle, e.g., "– What do ghosts have for dessert? – Icecream"), (ii) absurd humour (riddle based on incongruous humour, e.g., "– Have you heard about the restaurants on the moon? – Great food, but no atmosphere."), (iii) serious riddles (challenging questions about well-known quotations of writers, e.g., "– Who wrote the article 'J'accuse' ? – The answer is: Émile Zola."), and (iv) culinary riddle (questions about idioms related to food or cooking, e.g., "– What expression containing a vegetable means 'to be identical to someone'? – The answer is: 'be like two peas in a pod'.").

These riddle categories make it possible to observe a variety of reactions ranging from amusement/disinterest to more complex reactions, such as ones related to an unexpected intellectual challenge coming from the Nao robot.

2.2. Culinary challenge

The JOKER system aims at keeping the interaction entertaining also by initiating culinary-related challenges. At the moment, we have implemented the "discover my favourite dish" challenge. It consists in the robot asking the participant to guess a recipe name. The participant is expected to suggest recipes, or to ask culinary-related questions (e.g., about ingredients). The system evaluates the human contributions and reacts accordingly by including stimulating food-related interventions (e.g., "You are going to find the solution. It is as easy as pie!", "This recipe is not my cup of tea.").

2.3. Punctual interventions and teasing

The JOKER system provokes humorous situations by producing unexpected and judicious dialogue contributions. They take the form of food-related puns, funny short stories, well-known idiomatic expressions, or laughter. The choice of the content and of the timing of this kind of contribution is done by taking into account the human participant profile (emotional state, attitude towards the robot), and the dialogue history and context (e.g., challenge in progress, after a riddle).

The JOKER system also produces positive or negative comments about the participant or about itself. They mainly take

---

Abstract

We present automatic systems that implement multimodal social dialogues involving humour with the humanoid robot Nao for the 16th Interspeech conference. Humorous capabilities of the systems are based on three main techniques: riddles, challenging the human participant, and punctual interventions. The presented prototypes will automatically record and analyse audio and video streams to provide a real-time feedback. Using these systems, we expect to observe rich and varied reactions from international English-speaking volunteers to humorous stimuli triggered by Nao.

Index Terms: human-robot social interaction, humour, feedback.

1. Introduction

Humour is a strategy in social interaction that can help to form a stronger bond with each other, address sensitive issues, relax a situation, neutralise conflicts, overcome failures, put things into perspective or be more creative. From our perspective, some of the humour mechanisms can fruitfully be implemented in Human-Machine interaction. For instance, it could be used to overcome failure situations or to relax a situation.

For the 2015 Interspeech Conference, we would like to present the first prototype of the JOKER system resulting from efforts undertaken as part of the JOKER project[1] to enrich human-robot social dialogues with humour mechanisms. Our previous data collections were done using a WoZ (Wizard of Oz) system like in [1]. The automatic systems to be presented during the Show and Tell event have already shown promising results in realistic situations. They have actually been successfully used to collect about 8.25 hours of audio and video data with 37 French-speaking participants. An appropriate English-language version of these systems will be used for this demonstration.

2. Humour capabilities

Humour is a complex mechanism, built on a shared cultural referential between the teller and the listener. One crucial element in effective jokes is the *building and breaking* of tension: a story becomes funny when it takes an unexpected turn. Three main techniques will be used to generate reactions to humorous interventions from the robot: riddle, challenge and punctual interventions.

---

[1]See http://www.chistera.eu/projects/joker

place after revealing the answer of a riddle. They consist in: (i) positive comments about the participant (congratulations or encouragement, e.g., "You are very smart, these riddles are not easy at all."), (ii) negative comments about the human (gentle critics and teasing about how simple the question was or why the participant was not able to answer it, e.g., "A child could answer that!"), (iii) positive comments about itself (self-enhancing sentences, e.g., "My chipsets are much more effective than a traditional brain!"), and (iv) negative comments about itself (self-depreciating sentences, e.g., "I'm not very strong, look at my muscles.").

3. JOKER prototype

3.1. Architecture

The JOKER system makes use of paralinguistic and linguistic cues both in perception and generation, in the context of a social dialogue involving humour, empathy, compassion and other informal socially-oriented behaviour. Its architecture is depicted in figure 1.
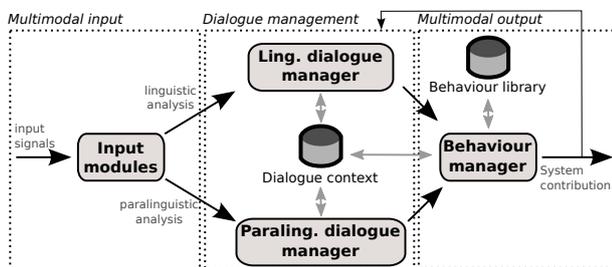

Figure 1: Architecture of the JOKER system

This system deals with multimodal input (audio and video). It includes the detection of paralinguistic cues (e.g., emotion detection system, laughter detection, speaker identification) [2] and linguistic cues (via an automatic speech recognition (ASR) system) [3]. Vision cues detection using kinect2 will be added later by partners of the project. Paralinguistic and linguistic cues are exploited by two specific dialogue managers that update a *shared* dialogue context (including dialogue history, dynamic user profile and affective interaction monitoring). These dialogue managers take advantages of the previously described humour strategies in response to various stimuli. The paralinguistic part manages the behaviour of the system in response to emotional and affect bursts stimuli. As for the linguistic one, it deals with lexical and semantic cues to provide an adequate response of the system. Eventually, the behaviour manager orchestrates contributions from both dialogue managers, and realises the output of the system through the Nao robot using one or more of these elements: speech, laughter, movements and eye colour variations.

3.2. Systems

Two automatic versions of the system will be used for this demonstration. They both explore essential aspects of the JOKER system in terms of architecture and of humour strategies. Preliminary versions of these systems have already been tested on a French-language data collection. The first automatic system focuses on the paralinguistic aspect. It involves a social interaction dialogue that adapts riddles telling to the user's profile. It implements an audio-based emotion detection module [2], vi-

sual identification of the speaker, a dynamic user model [4], and a finite-state based dialogue manager. The second automatic system focuses on the linguistic aspect of the system in the context of the "discover my favourite dish" challenge. It is based on the RITEL system, an open-domain dialogue system [5] and features an ASR, a question-answering system adapted to the culinary challenge, a natural language generation system, and a database of recipes and ingredients automatically crawled from the web. Tests indicate an average interaction duration of 2 up to 3 minutes for each version.

3.3. Sensors

We are naturally interested in speech in the context of the Interspeech Conference; audio streams will be captured during the experiment. Given the context of the recording (a conference room), we will use a directional microphone, to capture mostly participants' voice and limit noise. We are also considering affective facial expressions (e.g., smile, laughter) and speaker identification. To that purpose, we will make use of an off-the-shelf webcam to record a frontal video of volunteers.

4. Expected results

This event is an ideal moment to present the first results of the JOKER project. We will demonstrate two automatic systems displaying humour communication skills in the context of social Human-Robot interaction. The international venue of the Interspeech Conference will notably give us the opportunity to observe multicultural reactions to humorous stimuli.

This demonstration is an important opportunity to confront our first prototypes of the system with international English-speaking participants. To conclude, this demonstration will provide us with useful insights and feedbacks that will be valuable in the future data collection planned in the context of the JOKER project.

5. Acknowledgements

6. References

[1] M. Sehili, F. Yang, V. Leynaert, and L. Devillers, "A corpus of social interaction between nao and elderly people," in *5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data (ES3LOD2014). LREC*, 2014.

[2] L. Devillers, M. Tahon, M. A. Sehili, and A. Delaborde, "Inference of human beings' emotional states from speech in human–robot interactions," *International Journal of Social Robotics*, pp. 1–13, 2015.

[3] A. Rousseau, L. Barrault, P. Deléglise, Y. Estève, H. Schwenk, S. Bennacef, A. Muscariello, and S. Vanni, "The lium english-to-french spoken language translation system and the vecsys/lium automatic speech recognition system for italian language for iwslt 2014," in *International Workshop on Spoken Language Translation (IWSLT)*, Lake Tahoe (USA), 2014.

[4] A. Delaborde and L. Devillers, "Use of nonverbal speech cues in social interaction between human and robot: emotional and interactional markers," in *Proceedings of the 3rd international workshop on Affective interaction in natural environments*. ACM, 2010, pp. 75–80.

[5] B. van Schooten, S. Rosset, O. Galibert, A. Max, R. op den Akker, and G. Illouz, "Handling speech input in the Ritel QA dialogue system," in *InterSpeech'07*, Antwerp, Belgium, 2007.