Multimodal Data Collection of Human-Robot Humorous Interactions in the JOKER Project

CONFERENCE PAPER · SEPTEMBER 2015

18 AUTHORS, INCLUDING:

Laurence Devillers
Computer Sciences Laboratory for Mechanic…
126 PUBLICATIONS **1,738** CITATIONS

SEE PROFILE

Sophie Rosset
Computer Sciences Laboratory for Mechanic…
110 PUBLICATIONS **739** CITATIONS

SEE PROFILE

Amine Sehili
Computer Sciences Laboratory for Mechanic…
14 PUBLICATIONS **62** CITATIONS

SEE PROFILE

Yannick Estève
Université du Maine
71 PUBLICATIONS **314** CITATIONS

SEE PROFILE

Multimodal Data Collection of Human-Robot Humorous Interactions in the JOKER Project

Laurence Devillers* **, Sophie Rosset*, Guillaume Dubuisson Duplessis*, Mohamed A. Sehili*,
Lucile Béchade*‖, Agnès Delaborde*, Clément Gossart*, Vincent Letard*‖, Fan Yang*‖, Yücel Yemez†,
Bekir B. Türker†, Metin Sezgin†, Kévin El Haddad‡, Stéphane Dupont‡, Daniel Luzzati§, Yannick Estève§,
Emer Gilmartin¶ and Nick Campbell¶

* LIMSI-CNRS, Paris, France
** Université Paris 4, ‖ Université Paris Sud
† Koç University, Istanbul, Turkey
‡ UMONS, Mons, Belgium
§ LIUM, Le Mans, France
¶ TCD, Dublin, Ireland

***Abstract*—Thanks to a remarkably great ability to show amusement and engagement, laughter is one of the most important social markers in human interactions. Laughing together can actually help to set up a positive atmosphere and favors the creation of new relationships. This paper presents a data collection of social interaction dialogs involving humor between a human participant and a robot. In this work, interaction scenarios have been designed in order to study social markers such as laughter. They have been implemented within two automatic systems developed in the JOKER project: a social dialog system using paralinguistic cues and a task-based dialog system using linguistic content. One of the major contributions of this work is to provide a context to study human laughter produced during a human-robot interaction. The collected data will be used to build a generic intelligent user interface which provides a multimodal dialog system with social communication skills including humor and other informal socially oriented behaviors. This system will emphasize the fusion of verbal and non-verbal channels for emotional and social behavior perception, interaction and generation capabilities.**

***Index Terms*—Multimodal Data, Human-Robot Interaction, Humorous Robot**

I. INTRODUCTION

Our aim is to study social signals such as laughter occurring *during* multimodal social dialogs that involve humor between a human and a robot. This study is part of a longer term goal to build a generic intelligent user interface that provides a multimodal dialog system with social communication skills including humor and other informal socially oriented behaviors. To achieve these purposes, we conducted a data collection campaign with two automatic systems developed at LIMSI. This data collection implements interaction scenarios designed in order to study social markers such as laughter. This will confer us the opportunity to analyze the reaction of people with respect to the humor techniques performed by the robot.

Audiovisual databases containing laughter already exist (see [1] for a detailed review). They have been collected in the context of Human-Human interaction [2], Human-Agent interaction [3], multi-party meetings [4], [5] and laughter elicitation [1]. The data collection presented in this paper differs from these efforts by collecting social signals in the context of dyadic Human-Robot humorous interaction. Spontaneous reactions of human participants have been collected via audio and video channels together with Kinect2 data.

Laughter in social interaction has been subject to study lately (e.g., see the ILHAIRE project[1] [6]). In particular, studies in Human-Robot social interaction have focused on the social effect of laughter produced by a robot [7]. The originality of our work is to provide a very good material to study laughter produced by a human interacting with a robot in a humor context as well as to use the humor of the robot in interaction to create amusement and to engage the human in a relationship with the robot.

Humor has an important role in social relationships: from attracting in a first meeting to a long-term commitment. In new relationships, humor can be an effective way not only to attract each other, but also to overcome any awkwardness or embarrassment that arises while we get to know each other. In longer-term relationships, humor keeps a certain level of excitement, of shared pleasure. The shared humor creates a sense of intimacy and quality that is one of the founding principles of strong relationships. Laughing together allows to create a positive relationship. Humor can also help to overcome conflicts or disagreements. In summary, humor in human interactions can help to: form a stronger bond with each other, address sensitive issues, relax a situation, neutralize conflict, overcome failures, put things into perspective, be more creative.

Some of the humor mechanisms can be implemented on a machine. Humor could be used in failure situations (human or machine), to relax a situation, to help overcome these failures. Implementing a humorous behavior into human-machine interactions can take advantage of the potential of humor in establishing social relationships. It fosters a positive and friendly environment that facilitates the interaction and can increase cooperation with the system [8]. Humorous comments from a computer encourage people to make more sociable comments, and to joke back more [9]. It can also enhance the flow of interaction especially in conflicting and ambiguous situations [10]. But conversational joking has been described

[1] ILHAIRE project: http://www.ilhaire.eu/

both as conducive and aggressive to rapport, provoking agreement of the hearer as well as shock or indignation [11].

Evaluation of the relationship between user and system should be multi-faceted. A promising paradigm for evaluation of human-machine relationship is based on Communication Accommodation Theory, viewing relationship as the set of collaborative tasks participants are willing to engage in at any given time and thus as the level to which participants will accommodate each other [12]. It is proposed to use a similar methodology for evaluation in the JOKER project, using verbal, nonverbal behavior and contextual information. Automatic detection of engagement [13] will be tested in this project by combining several measures such as the number of laughter or of smiles, the positive emotion in the dialog. In this paper, we evaluate the interaction between subjects and the robot using questionnaires.

In this paper, section II presents work related to the design of multimodal social dialogue systems involving social signals such as laughter. Section III concerns the JOKER system and presents the multimodal data collection involving *automatic dialog systems* exhibiting the following features: emotion detection, dynamic user profile, generation system, synthesis through the Nao robot (speech, laugh, movement), database of jokes and knowledge about the cuisine domains, humor strategies, a system-directed dialog manager based on paralinguistic cues, and a dialog manager based on an interactive question answering (QA) approach. Section IV is devoted to the collection of multimodal data (audio, video, Kinect2) with three socially-oriented behavior scenarios in French language. Section V describes the first data analyses involving laughter and satisfaction of dialogue participants. Section VII concludes this paper and presents perspectives.

II. RELATED WORK

There exist many user interfaces providing a multimodal dialog system involving social communication capabilities (e.g., Semaine [14], Herme [15]). In recent work, a small scale "chatty" robot controlled by a human operator successfully engaged members of the public in social talk consisting of a series of two or three-turn exchanges. These exchanges incorporate short friendly jibes, more traditional "call and respond" jokes or riddles, and naturalistic feedback expressions such as "oh" and "really", which, along with naturalistic timing provided by the robot operator, give the impression of copresence and sociability on the part of the system [16].

A multimodal social dialogue system requires robust detection of non-verbal language. Recent work have demonstrated such capabilities in terms of detection of emotions in audio [17], detection of laughter [18]–[20] and detection of affect bursts, over audio and visual channels [21], [22]. In addition to non-verbal language, a social system can benefit from robust module for verbal language such as automatic speech recognition systems (see, e.g., [23]).

Social interactions require social understanding (planning ahead and dealing with new circumstances, anticipating the mental state of another person). Recent work highlights user models that allow the system to adapt its behavior by taking into consideration the user's personality, the user's interactional behavior (liking, dominance, familiarity) and the user's profile (age range, social hierarchy and relation) [24].

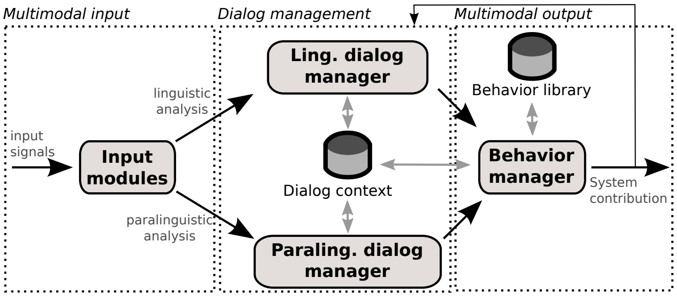

Figure 1. Architecture of the JOKER system.

Enhancing the repertoire of expressive and affective vocalizations that a social robot is able to generate is crucial for increasing its versatility and believability. Within the scope of humor and amusement, smile and laughter are indeed very relevant. They also affect speech production. Current research hence cover the generation of isolated laughter [25], as well as the generation of amused voice, colored by smile or laughter [26].

III. JOKER SYSTEM

A. Description

This work aims at building a generic intelligent user interface which provides a multimodal dialog system with social communication skills including humor and other informal socially oriented behaviors. This system will emphasize the fusion of verbal and non-verbal channels for emotional and social behavior perception, interaction and generation capabilities. It is also meant to run in real-time and to include a robust perception module (that perceives the user's voice, emotion, head orientation, facial expression and gaze), a rich social interaction module (to model the user and the context with long-term memories), an automatic speech recognition module, and a generation and synthesis module to maintain social engagement with the user. Laughter and social markers will particularly be used to build a dynamic user profile and to measure their engagement. Ultimately, this platform will be advantageously used to explore advanced dialogs involving complex social behaviors in order to create a long-term social relationship.

B. Architecture

The JOKER system aims at the fusion of paralinguistic and linguistic cues both in perception and generation, in the context of a social dialog involving humor. The envisioned architecture for the JOKER system is depicted in figure 1.

This system operates on multimodal inputs, and plans to take simultaneously into account audio, video and kinect2 data. Paralinguistic and linguistic cues are exploited by two specific dialog managers that update a *shared* dialog context.

This context contains key elements to interpret the communicative behavior of the human participant and to generate an appropriate reaction. First, it includes a dialog history that represents the current dialog structure in terms of lexical, semantic and pragmatic aspects (it contains, e.g., the dialog history). Next, it features a dynamic profile of the human dialog participant and an affective interaction monitoring module. Rich user models consider the user's personality (extroversion, optimism, self-confidence, and emotionality dimensions), the user's interactional behavior (liking, dominance, trust) and the user's profile (age range, social hierarchy and relation) to dynamically interpret multimodal cues on emotion and social dimensions during interactions.

Dialog managers take advantages of humor strategies (cf. section IV-A) in response to various stimuli. The paralinguistic one manages the behavior of the system in response to emotional and affect bursts stimuli. The linguistic one deals with lexical and semantic cues to provide an adequate response of the system.

Eventually, the behavior manager is in charge of the communicative behavior of the JOKER system. First, it orchestrates the contributions coming from both dialog managers by selecting the action to perform. Then, it generates a multimodal contribution of the system in terms of speech, affect burst, movements and eye color. Eventually, it activates a multimodal action scheduler that realizes the output of the system through the Nao robot.

C. Automatic Data Collection Systems

Three systems have been used for the data collection (one fully automatic system, one semi-automatic and one Wizard of Oz). They have been designed to explore essential and complementary aspects of the JOKER system in terms of paralinguistic/linguistic inputs and humor strategies. These systems were developed to collect rich and varied multimodal data. In this social interaction context involving humor, we are chiefly interested in the contextual occurrence of laughter.

1) Paralinguistic System: The first system focuses on the paralinguistic aspect of the JOKER system. It is fully automatic and features an emotion detection module based on audio [17], a dynamic user model [24], and a finite-state based dialogue manager. It involves a social interaction dialog that adapts the telling of riddles to some aspects of the user model. This model consists of two representations: interactional (user's attitude towards the robot) and emotional (user's affective tendencies in the course of the interaction). Its dimensions are automatically updated thanks to a decision system, which relies on data transmitted by the emotion detection module and the dialogue manager. For example, the overall expressiveness of the user is computed according to the strength of the emotions expressed by the user, and the duration of his or her speech. The update of the dimensions is based on expert boolean and fuzzy rules [24], which drive the computation of each dimension's score. In the JOKER system, the interactional dimensions of the profile endows the robot with a comprehension of the user's receptiveness to humor before selecting a behavior. Future works will also merge the emotional profile dimensions for the selection of the robot's behavior.

2) Linguistic System: The second system is semi-automatic and explores the linguistic aspect of the JOKER system in the context of the "discover my favorite dish" challenge offered to the dialogue participant (cf. section IV-A). It includes a question-answering system adapted to the culinary challenge similar to the open-domain dialogue system RITEL [27], a natural language generation system, and a database of recipes and ingredients automatically crawled from the web. Due to the poor availability of French resources to build an automatic speech recognition (ASR) system adapted to this task, speech recognition has been carried out manually. A human operator has typed utterances produced by the participant. Therefore, the system is automatic except for the speech recognition process. One goal of this data collection is to overcome the lack of such ASR.

3) Wizard of Oz: The third system is a Wizard of Oz dedicated to social dialog via the Nao robot, featuring all the humorous capabilities that we have currently designed (developed in section IV-A). It consists in a software with a graphic user interface remotely controlled by a human operator. It is configured by a predefined dialog tree that specifies the text utterances, gestures and laughter that can be selected to be executed by Nao. At each node, the operator chooses the next node of dialogue to visit according to the human dialogue participant's reaction.

IV. DATA COLLECTION PROTOCOL

A. Humor Capabilities

Three main techniques can intervene into our data collection scenarios in order to generate laughter (and other social signals) in reaction to humor: riddle, challenge and punctual interventions.

1) Riddles: The JOKER system keeps the social interaction entertaining by telling riddles. Its riddles follow a common structure. First, the system asks a question forming the riddle. Riddles are made so that the answer is not expected to be found. Then, the human dialog participant reacts to the riddle, for instance by suggesting an answer. Finally, the JOKER system provides the right answer and makes a positive or negative comment about the previous human contribution (cf. section IV-A3). The riddle database can be divided into four categories (examples are adapted in English): (i) social humor (socially acceptable riddle, e.g., "– Why did the tomato blush? – Because it saw the salad dressing."), (ii) absurd humor (riddle based on incongruous humor, e.g., "– How do you know there's been an elephant in the fridge? – Footprints in the butter."), (iii) serious riddles (challenging questions about well-known quotations of writers, e.g., "– Who wrote 'All the world's a stage and all the men and women are merely players'? – The answer is: William Shakespeare."), and (iv) culinary riddle (questions about idioms made on food or cooking, e.g., "– What expression about baked goods means the best ever? – The answer is: 'The greatest thing since sliced bread.'"). Our database contains approximately 20 items uniformly distributed over the categories.

2) Culinary challenges: The JOKER system keeps the interaction entertaining by initiating culinary-related challenges. At the moment, we have implemented the "discover my favorite dish" challenge. It consists in the robot asking the human dialog participant to guess a recipe name. The human 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.").

The main semantic domain chosen within this challenge is food and recipes. Different kinds of data were considered: many recipes in order to represent the domain (consisting of a title, ingredients, preparation and cooking times, and a description of steps to follow), and a list of ingredients along with information about their nature (e.g., meat, fish, fruit, vegetable, cheese). Database of recipes and ingredients was automatically constituted by crawling websites. It contains more than 63000 recipes and 2000 ingredients.

3) Punctual interventions and teasing: The JOKER system brings about humorous situations by producing unexpected and judicious dialog contributions in order to generate laughter. They take the form of food-related puns, funny short stories, well-known idiomatic expressions, or laughter. These contributions are selected by taking into account the human participant profile (emotional state, attitude towards the robot), and the dialog history and context (e.g., after a human contribution during a challenge, after revealing the solution of a riddle).

The French language has many metaphorical idioms involving food or cooking (e.g., in English, "As easy as apple pie", "It's a piece of cake"). We constituted a playful set of vocabulary made of metaphoric collocations on food (approx. 50 items). These metaphoric idioms make it possible to create puns related to food in order to extend the humorous capabilities of the robot.

The JOKER system also produces positive or negative comments about the participant or about itself. They mainly take place after revealing the answer of a riddle. They consist in: (i) positive comments about the human (congratulations or encouragement, e.g., "You're doing really well! Congratulations!"), (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.").

Finally, the JOKER system generates laughter at given times of the interaction: after telling a humorous contribution (e.g., after a riddle, a short story) and to alleviate the effect of negative comments about the human participant.

B. Conversational Scenarios

Each data collection system displays a subset (if not all) of the previously described humor capabilities.

System 1 (paralinguistic) implements a social dialog in which the system automatically adapts the telling of riddles to the dynamic user profile. The interaction starts with a greeting phase in which Nao presents itself. Next, the robot proposes the telling of a riddle adapted to the detected emotional state of the human. Then, the behavior of the system depends on the receptiveness of the human to the contributions of the robot. Positive reactions lead to more riddles and funny short stories, whereas repeated negative reactions drive the dialog to a rapid end. In the end, the system closes the interaction by drawing a conclusion about the perceived reactions from the human (e.g., "I am very glad you like humor produced by a robot."). This system features the following humor capabilities: riddles, positive and negative comments, funny short stories, and laughter.

System 2 (linguistic) interacts with the human dialogue participant in the context of the "discover Nao favorite dish" challenge. Interaction comes down to the following structure: (i) greeting phase, (ii) challenge, and (iii) closing phase. This system exhibits the culinary challenge, food-related puns and idioms capabilities.

Eventually, system 3 (Woz) displays all the humor categories by seamlessly combining the scenarios of the two first systems. It starts with the culinary challenge featured in system 2, and then continues with some riddles following a structure similar to the scenario of system 1. This system contains more than 230 specified sentences.

V. Corpus Description

A. Experimentation Setup

This experiment took place in the cafeteria of the LIMSI. Volunteers were seated facing the Nao robot (at around one meter from it). A webcam was placed in front of participant from a distance about one meter at theirs eyes level height. A Microsoft's Kinect2 was placed just under the webcam. Beside the webcam, a panoramic camera has been placed at 45 degrees on the right side of the participant from a distance about 3 meters to take a global vision of the experiment.

Volunteers interacted with the three systems in the following order: (1) system 1 (paralinguistic, automatic), (2) system 2 (linguistic, semi-automatic), and (3) system 3 (Woz). Between each system, participant were asked to fill satisfaction questionnaires. Before entering the social dialogue, system 1 was used to perform a pre-test as a game where the system asked the human participant to play some emotion (e.g., "Please, say something joyfully!"), and then told him what it recognized (e.g., "I have detected joy in your contribution."). This pretest was done to expose to the human the emotional detection capabilities of the system. A full interaction session lasted approximately 30min per participant.

All participants were volunteers working at the LIMSI laboratory, and French speaking. We recorded 37 participants (62%

Table I
DURATION OF THE COLLECTED DATA PER SYSTEMS. σ IS THE STANDARD DEVIATION.

	System 1 (Automatic)	System 2 (Semi-auto.)	System 3 (Woz)	Total
total	3h 12m 32s	1h 25m 20s	3h 20m 57s	7h 58m 50s
average	5m 12s	2m 18s	5m 25s	3m 14s
σ	26s	55s	1m 00s	1m 28s

male, 38% female). The average age of the participants is 35.1 (standard deviation: 11.97; min: 21; max: 62). Volunteers were asked to fill a personality questionnaire (OCEAN [28]) and the Sense of Humor Scale (SHS) questionnaire [29]. In this experiment, participants SHS score range from 72 to 145 (mean: 108.87, standard deviation : 22.38).

B. Collected Data

For each participant, three kinds of data have been collected: audio, video and kinect2.

1) Audio: A high-quality AKG Lavalier microphone has been used for audio data acquisition. The acquired data consist of audio tracks of 16kHz for the automatic part (recorded internally by the system) and 44.1kHz for the Woz and the semi-automatic system (recorded using Audacity). Each audio track represents one single interaction between a subject and the robot. Although the microphone was fixed on the subject's clothes, the relatively small distance between the two interlocutors allowed for a good capture of the robot's utterances, laughter and other sounds.

2) Video: A logitech HD Pro C920 Webcam has been used for facial and shoulder image recording. The experiment has been recorded using H.264 codex in 720p image quality with 30 frames per second and 16 kHz processed audio. The webcam image focused on the facial details which will be investigated for attention detection and facial expression recognition. Beside the webcam, a panoramic video has been recorded with a Sony HDR-CX410 digital video camera in 1080p image quality with 25 frames per second using H.264 codex and in 16 kHz processed audio from 5.1 channels build-in microphones.

3) Kinect2: Microsoft Kinect2 was used to capture facial features and upper-body gestures. Using Kinect2, we have recorded low level streams, namely, video and depths frames (rgb-d) at 30 frames per second and 16 kHz processed audio. Kinect proprietary software development kit also provides high level streams such as body skeletons, facial animations parameters and facial features which are also recorded and stored. Details of these features can be found at the Microsoft website [30].

4) Synthesis: Table I presents the recording durations for each system. All in all, we recorded approximately 8 hours of data for each kind of input (audio, video, and kinect2). Duration reported for the system 1 includes both the pre-test and the social dialogue. The social dialogue alone accounts for 1h 30min 5s of recording (average: 2min 26s; standard deviation: 14s). Unsurprisingly, an average duration of interaction with the system 3 is superior to the average durations with system 1 or system 2 (because of its inclusion of both riddles and culinary challenge).

VI. PRELIMINARY RESULTS

A. Examples from the Corpus

We now present some transcribed and translated examples from the collected corpus that show how the humor capabilities of the JOKER system provide a context to study the occurrence of different kinds of laughter. By no means should these examples be taken as an exhaustive list.

Naturally, we have observed *spontaneous laughter* generated by a joke that has genuinely been found funny, e.g. ("N" stands for "Nao", "P" for "Participant"):

- N: This reminds me of an anecdote: to fall asleep, a sheep can only count on itself.
- P: [laugh] Nice one! [laugh]

In this example, Nao has detected that the human participant likes its humor. As a consequence, it chooses to tell a joke (here, a funny short story). The human participant reacts with a spontaneous laughter and a comment showing his appreciation of the joke. Spontaneous laughter can also be triggered by an unexpected intervention from Nao such as a positive comment. We have observed such occurrences after comment like "You look radiant!".

We also observed *politeness laughter* in the context of a riddle that has not been found as funny as intended, e.g., after an absurd one:

- N: Why are there no more "mammoths"?
- P: *(pause)* euh *(pause)* I have to admit that I have no idea!
- N: Well, the answer was: because there are no more "papoths"! [laugh]
- P: [laugh] OK, thank you for the riddle. [laugh]

Laughter from the human participant is first triggered by mimicry (following the laughter generated by Nao), and then maintained by politeness as shown by the thanks.

Eventually, we have also observed *mitigation laughter* in the context of a negative comment, which aims at alleviating the negative content of an utterance, e.g.:

- N: Even my little sister would have succeeded! [laugh]
- P: Well... good thing I am not your sister! [laugh]

Here, the human participant reacts to a negative comment from Nao by another negative comment, and ends it with a mitigation laughter.

B. Study of Satisfaction Questionnaires

Satisfaction questionnaires consisted of closed-ended question about the system, the interaction and the human participant. They were filled by the participants immediately after an interaction with a data collection system. Participants thus filled three satisfaction questionnaire (one for each system).

Participants were asked to answer a question about the nature of the system they had just been interacting with:

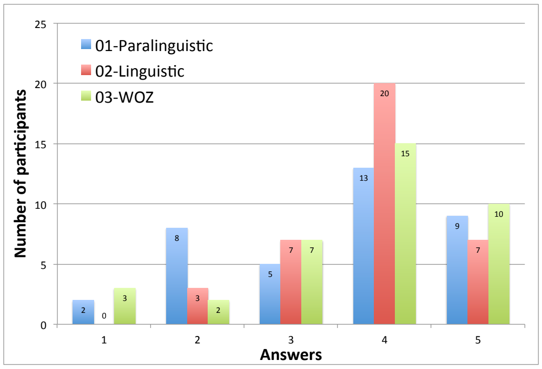

Figure 2. Distribution of the answers of the participants to the question "Did you have any desire to talk to the robot?" over the three systems. From 1 ("Strongly disagree") to 5 ("Strongly agree").

"According to you, was the system *automatic*, *semi-automatic* or *manually directed*?". For each system, the majority of participants identified its right nature: 60% of participants have considered the first system to be fully automatic, 50% have identified the second system to be semi-automatic, and 48% of participants have selected a directed system for the Woz. These results can be explained by the fact that the participants are used to interact with such systems.

Next questions relate to an assessment of the satisfaction of the interaction with the robot, amusement on the interaction, desire to talk to the robot and the adaptation of the system to the participant using a Likert scale of 5 points. We report here the results for each system in a triplet $(\text{system } 1, \text{system } 2, \text{system } 3)$ (from "1: Strongly disagree" to "5: Strongly agree").

- "Did you like the interaction with the robot?" (mode: $(4, 4, 4)$; median:$(4, 4, 4)$)
- "Were you amused?" (mode: $(4, 5, 4)$; median: $(4, 4, 4)$)
- "Did you have any desire to talk to the robot?" (mode: $(4, 4, 4)$; median:$(4, 4, 4)$)
- "Did the robot adapt itself to you?" (mode: $(4, 3, 4)$; median: $(3, 3, 4)$)

Overall, for the three systems, participants have considered themselves amused and satisfied by the interactions. They globally think that systems have adaptation capabilities. Participants have felt a desire to talk to the robot as shown by the detailed results presented on figure 2. Notably, this desire kept increasing over the interactions: the overall distribution of agreements (> 3) increases from session 1 and 2, and then it plateaus between session 2 and 3.

Participants were asked to evaluate the attitude of the robot using a Likert scale of 5 points: "According to you, the attitude of the robot was ..."

- ...friendly (from "1: Friendly" to "5: Hostile"): mode:$(1, 1, 1)$; median: $(1, 2, 1)$.
- ...submissive (from "1: Submissive" to "5: Dominant"): mode:$(4, 3, 3)$; median: $(4, 3, 3)$.

Participants have globally considered the robot as being very friendly over the sessions. Further study should be considered to clearly distinguish what part of this result can be attributed to Nao in itself (shape, voice, etc.) and to the interaction. On the other hand, the robot have been mainly evaluated as neither submissive nor hostile. The fact that the robot lead the interaction obviously contributed to not being viewed as submissive. In particular, the robot was not found to be hostile despite its teasing interventions. All in all, these results seems consistent with the fact that participants reported being satisfied with the interactions and generally amused.

Eventually, participants were asked to assess their own attitude. For the three systems, participants have clearly disagreed with the negative adjectives ("hurt", "embarrassed") and have agreed with the positive ones ("amused", "confident").

VII. Conclusion and Future Work

In this paper, we have described a multimodal data collection involving two automatic systems in the context of social dialog exhibiting humor between a human and a robot. We collected 8 hours of audio, video and kinect2 data of social interaction. Interactions with the systems have largely been rated by the participants as satisfying and amusing. We have presented and implemented a rich set of humorous techniques (riddles, punctual interventions, soft teasing, culinary challenges) deployed by the systems to elicit reactions such as laughter from the human dialog participant. We have shown how these humor strategies provide a *context* to study laughter occurring *in* social interaction dialogs. We have presented an architecture of a future generic intelligent user interface providing a multimodal dialog system with social communication skills including humor and other informal socially oriented behaviors. Design of this system will take advantage of the collected data to build robust perception module, a social interaction module modeling user and context with long-term memories, an automatic speech recognition system, and a generation and synthesis module for maintaining social engagement with the user.

This work raises interesting perspectives. First, it consists in performing a quantitative study of laughter contextualized in the presented framework. As stated in the introduction, we more generally intend to evaluate the user-robot relationship as well as user engagement by combining verbal and nonverbal behaviour and contextual information. Next, we are currently planning two new data collection with two well-defined goals. One will take place with an adapted version of the presented system in English. We hope to use the collected data to perform a multi-cultural comparative study of social interaction involving humor with a robot. The other data collection will involve the same participants as the data collection presented in this paper, and aims at studying the impact of a long-term relationship between the human participant and a robot. To that end, we are going to integrate a memory into the JOKER system including static knowledge about the participant (e.g., name, age) and carefully selected significant events from the previous interaction (e.g., an especially appreciated joke type).

ACKNOWLEDGMENT

The described work has been sponsored by ERA-Net CHIST-ERA (http://www.chistera.eu/), the "Agence Nationale pour la Recherche" (ANR, France), the "Fonds National de la Recherche Scientifique" (FNRS, Belgium), the "The Scientific and Technological Research Council of Turkey" (TUBITAK, Turkey, grant number 113E324) and the "Irish Research Council" (IRC, Ireland).

REFERENCES

[1] S. Petridis, B. Martinez, and M. Pantic, "The MAHNOB laughter database," *Image and Vision Computing*, vol. 31, no. 2, pp. 186–202, 2013.

[2] B. Schuller, R. Müller, F. Eyben, J. Gast, B. Hörnler, M. Wöllmer, G. Rigoll, A. Höthker, and H. Konosu, "Being bored? Recognising natural interest by extensive audiovisual integration for real-life application," *Image and Vision Computing*, vol. 27, no. 12, pp. 1760–1774, 2009.

[3] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder, "The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 5–17, 2012.

[4] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, and others, "The AMI meeting corpus: A pre-announcement," in *Machine learning for multimodal interaction*. Springer, 2006, pp. 28–39.

[5] N. Campbell, "Tools and resources for visualising conversational-speech interaction," in *Multimodal corpora*. Springer, 2009, pp. 176–188.

[6] S. Dupont, H. Çakmak, W. Curran, T. Dutoit, J. Hofmann, G. McKeown, O. Pietquin, T. Platt, W. Ruch, and J. Urbain, "Laughter research: a review of the ilhaire project," in *Socially Believable Behaving Systems – The Quest for Equipping Machines with Human-Level Automaton Intelligence*. Springer Series on Intelligent Systems Reference Library, 2015 (to be published).

[7] C. Becker-Asano and H. Ishiguro, "Laughter in social robotics-no laughing matter," in *Intl. Workshop on Social Intelligence Design*, 2009, pp. 287–300.

[8] A. Nijholt, "Conversational agents and the construction of humorous acts," in *Wiley Series in Agent Technology*, T. Nishida, Ed. Chichester, UK: John Wiley & Sons, Ltd, Nov. 2007, pp. 19–47.

[9] J. Morkes, H. Kernal, and C. Nass, "Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of srct theory." *Human- Computer Interaction*, vol. 14(4), pp. 395–435, 1999.

[10] P. Kulms, S. Kopp, and N. C. Kramer, "Let's be serious and have a laugh: Can humor a support cooperation with a virtual agent?" in *Intelligent Virtual Agents*, T. Bickmore, S. Marsella, and C. Sidner, Eds., vol. volume 8637. Springer International Publishing, 2014, pp. 250–259.

[11] N. Norrick, "Issues in conversational joking," *Journal of Pragmatics*, vol. 35, pp. pp. 1333–1359, 2003.

[12] T. Bickmore and D. Schulman, "Empirical validation of an accommodation theory-based model of user-agent relationship," in *Intelligent Virtual Agents*. Springer, 2012, pp. 390–403.

[13] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[14] M. Schroder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. Ter Maat, G. McKeown, S. Pammi, M. Pantic *et al.*, "Building autonomous sensitive artificial listeners," *Affective Computing, IEEE Transactions on*, vol. 3, no. 2, pp. 165–183, 2012.

[15] J. Han, E. Gilmartin, and N. Campbell, "Herme, yet another interactive conversational robot," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013, pp. 711–712.

[16] E. Gilmartin and N. Campbell, "More than just words: Building a chatty robot," in *Natural Interaction with Robots, Knowbots and Smartphones*, J. Mariani, S. Rosset, M. Garnier-Rizet, and L. Devillers, Eds. Springer New York, 2014, pp. 179–185.

[17] 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.

[18] M. T. Knox and N. Mirghafori, "Automatic laughter detection using neural networks." in *INTERSPEECH*, 2007, pp. 2973–2976.

[19] K. P. Truong and D. A. Van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Communication*, vol. 49, no. 2, pp. 144–158, 2007.

[20] S. Petridis and M. Pantic, "Audiovisual discrimination between speech and laughter: Why and when visual information might help," *Multimedia, IEEE Transactions on*, vol. 13, no. 2, pp. 216–234, 2011.

[21] B. B. Turker, S. Marzban, E. Erzin, Y. Yemez, and T. M. Sezgin, "Affect burst recognition using multi-modal cues," in *Signal Processing and Communications Applications Conference (SIU), 2014 22nd*. IEEE, 2014, pp. 1608–1611.

[22] B. B. Turker, S. Marzban, M. T. Sezgin, Y. Yemez, and E. Erzin, "Affect burst detection using multi-modal cues," in *Signal Processing and Communications Applications Conference (SIU), 2015 23th*. IEEE, 2015, pp. 1006–1009.

[23] A. Rousseau, G. Boulianne, P. Deléglise, Y. Estève, V. Gupta, and S. Meignier, "Lium and crim asr system combination for the repere evaluation campaign," in *Text, Speech and Dialogue*. Springer, 2014, pp. 441–448.

[24] 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*, ser. AFFINE '10. New York, NY, USA: ACM, 2010, pp. 75–80.

[25] J. Urbain, H. Çakmak, A. Charlier, M. Denti, T. Dutoit, and S. Dupont, "Arousal-driven statistical parametric synthesis of laughter," in *Selected Topics in Signal Processing*, vol. 8, no. 2. IEEE, 2014, pp. 273–284.

[26] K. El Haddad, S. Dupont, J. Urbain, and T. Dutoit, "Speech-laughs: an hmm-based approach for amused speech synthesis," *Trans. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015.

[27] 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.

[28] R. R. McCrae and O. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, pp. 175–215, 1992.

[29] P. McGhee, *The laughter remedy. Health, healing and the amuse system.*, I. Kendall/Hunt., Ed. Dubuque, 1996.

[30] Microsoft, "Microsoft kinect for windows features [online]," http://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx, 2015, accessed: 2015-04-21.