



JOKE and Empathy of a Robot/ECA: Towards social and affective relations with a robot

Seminar CHIST-ERA Bern: April 28 2016

(call IUI 2012)

<http://www.chistera.eu/projects/joker>

Sophie Rosset

Scientific background

General objective: making a robot more socially competent which implies some capacities:

- visual interpretation (eyes tracking, face, gesture)
- speech recognition
- affective and social dimensions in spoken interaction
- dialogue and interaction management
- natural language generation
- speech synthesis

→ Fields covered by the partners

Main Challenges

- Humor in human-machine interaction
 - Humor can trigger surprise, amusement, or irritation if it doesn't match the user's expectations.
 - Social interactions require social intelligence and understanding for dealing with new circumstances by anticipating the mental state of the speaker or addressee.
- JOKER project explores:
 - two social behaviors: expressing empathy and exchanging chat with the interlocutor as a way to build a deeper relationship,
 - robust perception modules: Fusion between paralinguistic and linguistic cues for understanding speaker intentions,
 - real-time reactions of the robot/ECAs,
 - long-term memories (user model, anticipation).

Partners and their complementarities



affective and social dimensions in spoken interaction, emotion detection, dialogue, generation, user models, IHRobot



social interaction, multimodal interaction, collection of data, affect bursts detection and generation



visual interpretation (eyes tracking, face, gesture), user detection using visual cues, temporal model of gesture, gaze and speech



speech recognition (Kaldi/Sphinx), models of humor



speech synthesis, affect bursts detection and generation (laugh, breath, sigh, throat, etc.)

Main Scientific Results

- Objectives of WP2 *Domain and Databases of interactive speech*
 - Define the task
 - Collect data in two languages: French and English
- Results
 - Interaction and dialogic corpus collection
 - Corpus collected in English in 2016, mostly comparable in terms of types of systems and tasks
 - Corpus collected with the UCAR chatbot system
 - Corpus collected in French with elderly people (audio and video data)
 - Audiovisual data collection:
 - 2 sessions of Taboo, interactive game recording with Kinectv2
 - Laughter data with video stimuli, Kinectv2
 - Corpus annotation
 - French corpus annotation on-going and common annotation schema
 - IEMOCAP (about 8 hours dyadic interactions), audiovisual, crosstalks, weak/strong laughters
 - 16 sessions of Joker dataset collected in LIMSI, audiovisual, laughters

Main Scientific Results

- Objectives of WP3 *Perception modules*
 - Develop the different perception modules
 - Non-verbal language: audio, visual, gestural, eye gaze ...
 - Verbal language: speech recognition and natural language understanding
 - Propose and develop approach for efficient multi-modal fusion
- Results
 - A fast large vocabulary ASR system for English and French languages adapted to the culinary task, run on “real time” on a laptop
 - Affect Bursts detection
 - Speaker independent laughter detection
 - Laughters under the effect of crosstalk analysis

Main Scientific Results

- Objectives of WP4 *Dialogue and Decision modules*
 - Cognitive communication model, based on:
 - Model of inference process built on different multi-modal information perceived
 - Model of the possible state transitions
- Results
 - Design and Implementation of multiplatform CARA spoken dialogue system and WoZ interface
 - Design and implementation of the English version of the UCAR system, an unsupervised linguistic social conversation system (Unsupervised Corpus-based Answer Retriever)
 - Design and Implementation of the Paralinguistic JOKER-LIMSI interaction system and WoZ interface

Main Scientific Results

- Objectives of WP5 *Generation and synthesis modules*
 - Generate multimodal outputs following the information provided by the dialog module
 - Speech and affect bursts synthesis
- Results
 - Unsupervised answers selection and adaptation module for the UCAR system
 - Speech-Laugh: HMMs models for laughter and breath intake sounds (occurring usually after laughter)
 - Multilevel Amused Smiled Speech Synthesis: HMMs models for neutral and smile speech, trained on two collected databases

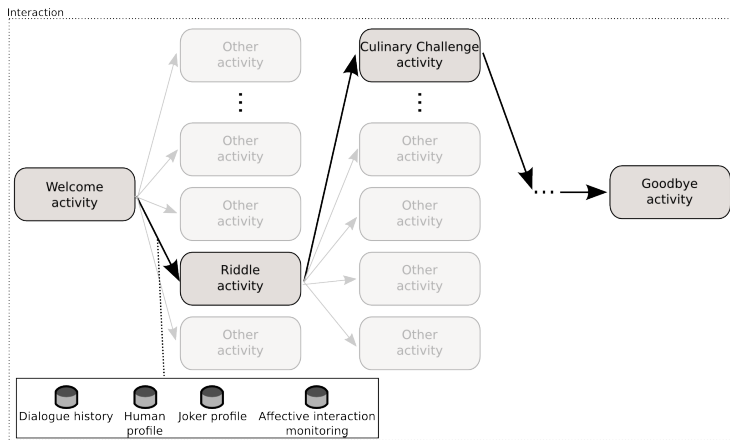
- Scientific events organization and participation
 - 3 workshops organized
 - 2 at LREC'16 (Just Talking – social talk among humans and machines, Ethica2 – Ethics in corpus collection, annotation and application)
 - MMSYM 2015 in Dublin
 - Demos
 - Interspeech'15, Show and Tell
 - LREC RE-WOCHAT Workshop, LIMSI was data provider, data generator and annotator
 - eNTERFACE 2015
 - Organized by UMONS
 - TCD participated with the project of adding social and joking capacity to an educational dialogue system in the Computer Aided Language Learning (CALL) domain

- Available corpora
 - UCAR corpus available at <https://ucar.limsi.fr>
- Publications
 - All: 2 international consortium publications
 - LIMSI: 2 national, 5 international workshops or conferences, 2 journal
 - LIUM: 3 international workshop and 1 book chapter
 - UMONS: 6 international conferences and workshops and 2 symposiums
 - KOC: 2 international conferences
 - TCD: 1 book chapter, 8 international conferences and workshops, 1 journal, 1 workshop proceedings

Dialogic Interaction: Combination

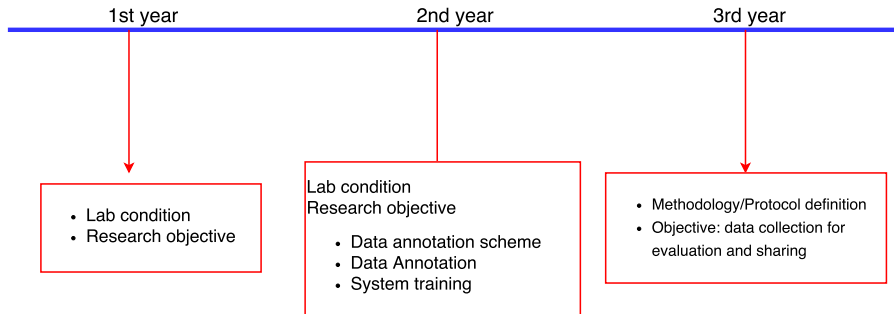
- Convergence/divergence:
 - mixture of both channels
 - influence between both (examples in the video)
- High level communicative process (semantic channel, task oriented)
- Low level communicative process (social interaction: attention, mimicry, laugh)

Dialogic Interaction: Adaptation



Joker Future Work

Database collection for evaluation



- Infrastructure needs: platform, data storage, annotation experts,
- Generally speaking: costs are underestimated in projects as such the Joker Project.