

CHIST-ERA Projects Seminar

JOKER♥Project

Laurence Devillers
Brussels, March 22-23, 2017
JOKER Ends September



- ❖ **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
- The template is flexible

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.)

Laugh with robots to live better with

- ❖ Several agents and robots have been used for doing these researches



❖ 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)



❖ Results: Audiovisual data collection

- ✓ Laughter data with video stimuli, Kinectv2
- ✓ Laughter Responsive Robot
- ✓ Scenario: Interactive quiz game (Question-answer)
- ✓ Evaluation : objective engagement measures (found significant difference in controlled experiments), questionnaires



❖ Results: Corpus annotation

- ✓ French corpus annotation on-going and common annotation schema
- ✓ Sessions of Joker dataset collected in LIMSI : audiovisual and multimodal annotation
- ✓ IEMOCAP (about 8 hours dyadic interactions), audiovisual, crosstalks, weak/strong laughters
- ✓ Work are going to be done to compare the French and English corpora (TCD, LIMSI, LIUM...)

❖ Objectives of WP3 - Perception modules

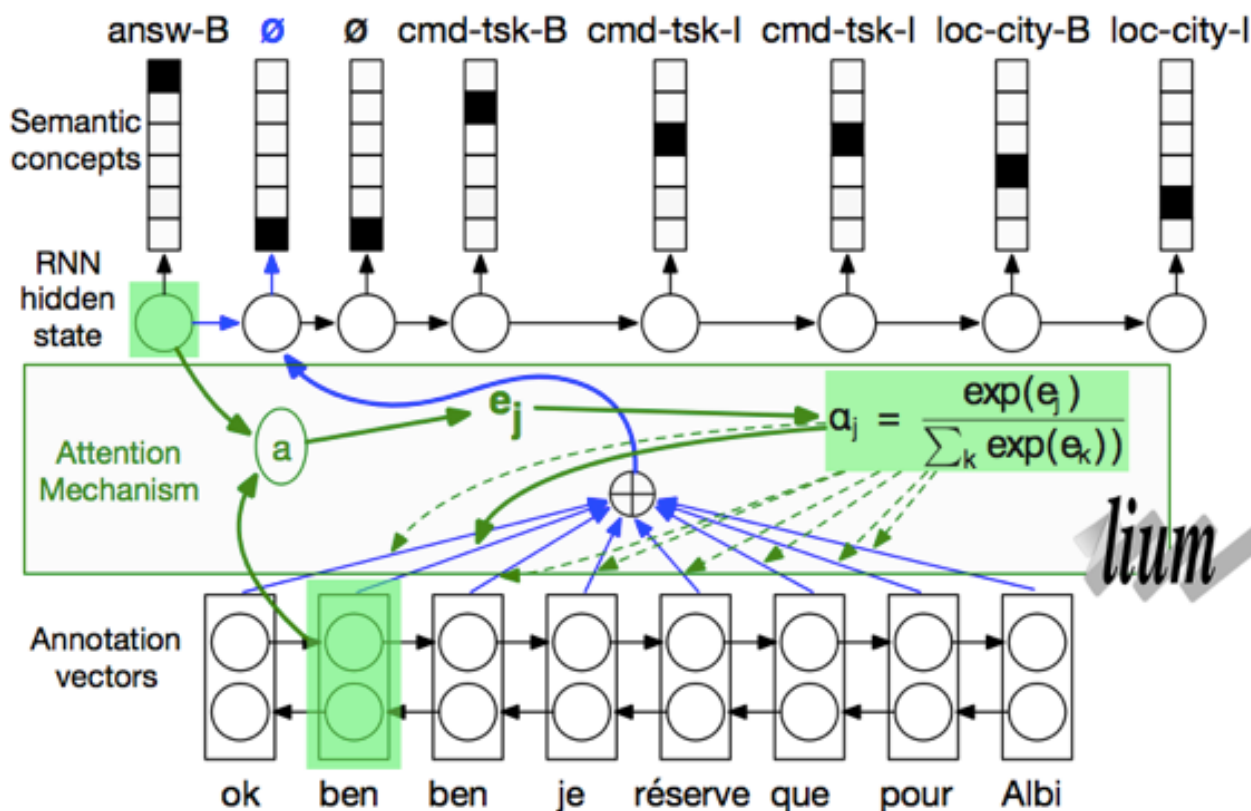
- ✓ Develop the different perception modules
- ✓ Non-verbal language: audio, visual, gestural, eye gaze ...
- ✓ Verbal language: speech recognition and natural language understanding
- ✓ Propose and develop approach for efficient multi-modal fusion

❖ Results:

- ✓ A fast large vocabulary ASR system for English and French languages
 - ✓ Two state-of-the-art ASR system were developed in the framework of JOKER
- ✓ In English : second position in the MGB 2015 evaluation campaign
 - ✓ No evaluation campaign for French
- ✓ For French: a real time large vocabulary embedded ASR system developed
 - ✓ Used to transcribe/translate speeches or PhD defense (also work for English)

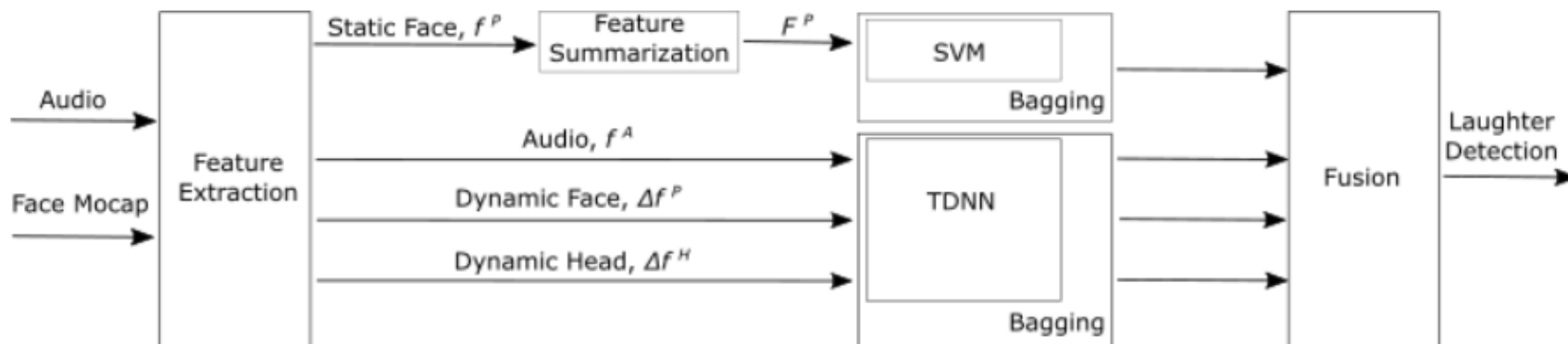
❖ Results: SLU: a neural approach

- ✓ A encoder/decoder system with attention mechanism has been developed in JOKER



❖ Results:

- ✓ Speaker independent laughter detection Laughters under the effect of crosstalk analysis
- ✓ Spectral and prosodic features from audio
 - ✓ Face, head motion capture
 - ✓ 8 hours of dyadic conversation data (IEMOCAP)



❖ 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
 - ✓ Improved and extended : Mavenised; Web-based; Documentation and packaging process
 - ✓ Investigating integration of stochastic modeling of dialogue phase Integrating various multimodal inputs (Mio sensor, engagement detection, presence detection)
 - ✓ Experimenting with different delivery form factors

❖ Results:

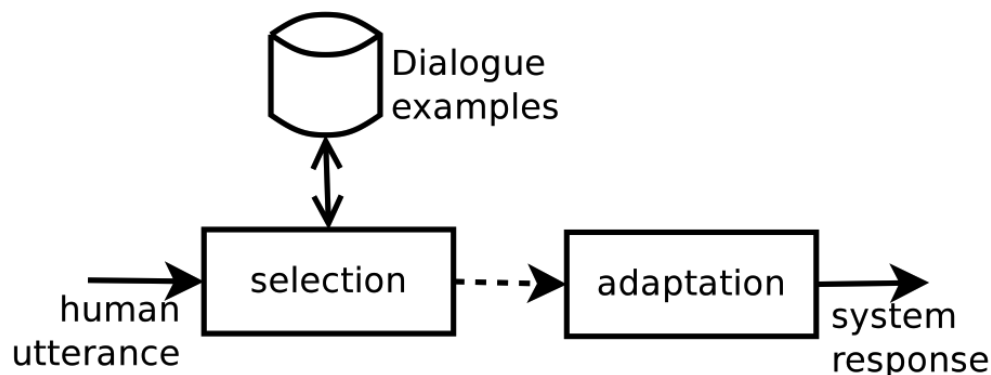
- ✓ Design and implementation of the English version of the UCAR system, an unsupervised linguistic social conversation system (Unsupervised Corpus-based Answer Retriever)
- ✓ Design and Implementation of the Paralinguistic JOKER-LIMSI interaction system and WoZ interface

❖ Objectives of WP5 - Generation and synthesis modules

- ✓ Generate multimodal outputs following the information provided by the dialog module
- ✓ Speech and affect bursts synthesis

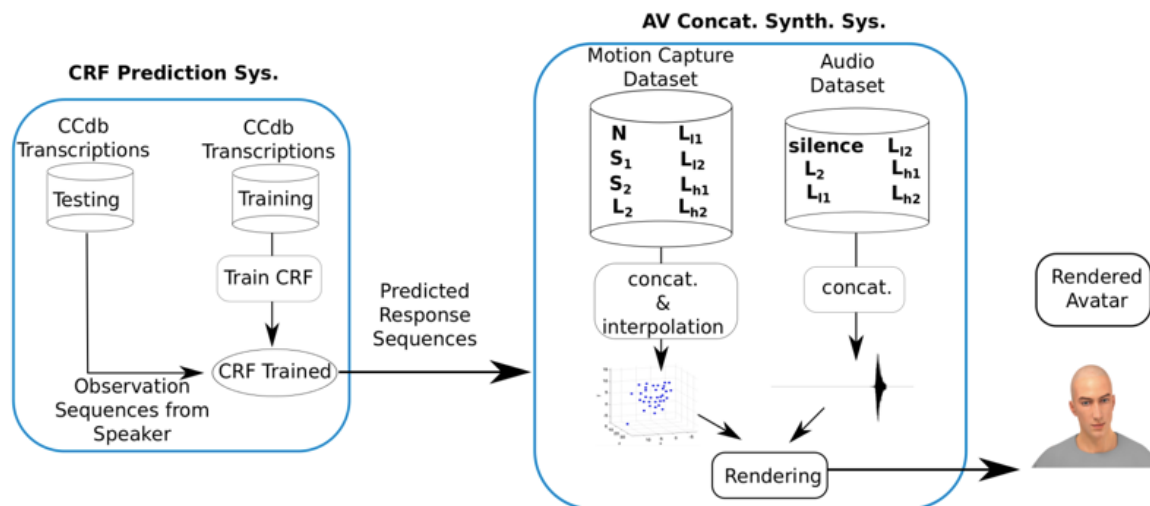
❖ Results:

- ✓ Unsupervised answers selection and adaptation module for the UCAR system



❖ Results:

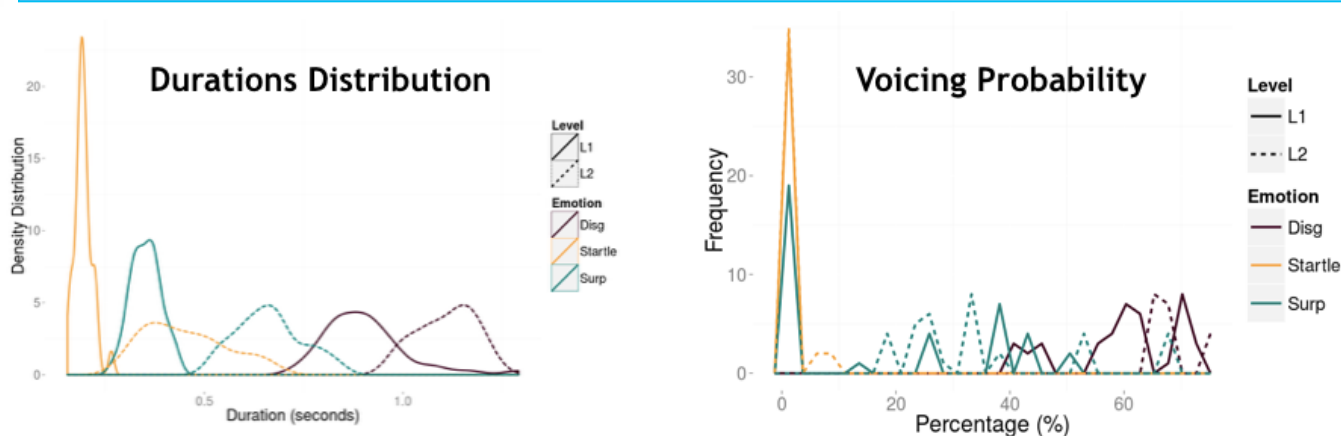
- ✓ Predicting and Generating AV Smiles and Laughs
 - ✓ Smiles and laughs as feedback to speaker
 - ✓ CRF predicts listener's states (smiles/laughs) as well as their duration, position and intensity level
 - ✓ Synthesis is AV concatenative-interpolation synthesis Synthesis can control duration and intensity of expressions



❖ Results:

- ✓ Speech Synthesis with Control of Amusement/Smiling Level
 - ✓ English and French amused/neutral speech database
 - ✓ Phoneme to wav and control of intensity level: french and english
 - ✓ GUI for multilingual multilevel smiled speech synthesis

Main Scientific Results - WP5



Affect Burst	Arousal Level	Instances
Disgust	Level 1	40
	Level 2	19
Surprise	Level 1	37
	Level 2	34
Fear	Level 1	34
	Level 2	39

❖ Results:

✓ Affect Burst Synthesis

- ✓ Motion Capture and audio database for 3 affect burts for: Disgust, surprise, startle
- ✓ HMM-based audio synthesis system

❖ 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
- ✓ GUI for multilingual multilevel smiled speech synthesis

❖ eINTERFACE 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

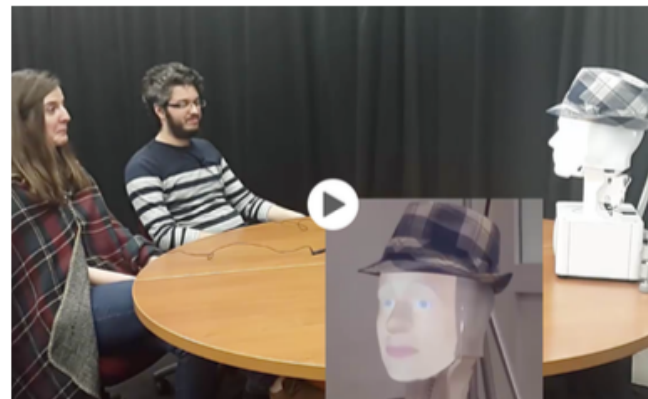
- ✓ UCAR corpus available at <https://ucar.limsi.fr>

❖ Publications

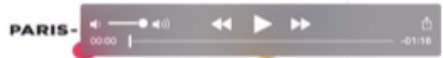
- ✓ All: 2 international consortium publications
- ✓ LIMSI: 2 national, 9 international workshops or conferences, 2 journal
- ✓ LIUM: 3 international workshop and 1 book chapter
- ✓ UMONS: 6 international conferences and workshops and 2 symposiums
- ✓ KOC: 2 international conferences
- ✓ TCD: 1 book chapter, 8 international conferences and workshops, 1 journal, 1 workshop proceedings

Database collection for evaluation

Data Collection - KOC team



Data Collection - LIMSI team

 **JOKER♥ Project**
Data Collections and Challenges (LIMSI-CNRS)
Feb 2016 - Feb 2017
<http://www.chist-era.eu/projects/joker>

Contact: laurence.devillers@limsi.fr 



Questions ?