# EASA Project
# Environment-Aware Social Agent

Hüseyin Çakmak, Kevin El Haddad, Nicolas Riche, Julien Leroy

Enhancing the human-machine interaction by adding emotions to the machine's way of expression, is one of the main topics of current research. This would improve the interaction relying on the assumption that more human-like the machine behavior is, more comfortable the interaction with it will be. This project proposes to create an environment aware emotional avatar. This avatar will be placed in an experimental framework which is described as follows: Participants will be interacting with each other in a limited room, in which the avatar will be present as a reactive spectator. The avatar should be able to recognize when different scenario events (specified scenarios described later) occur and react in an affective way with respect to each occurrence. It will react through different affective expressions and speech sentences. The impact that, adding an emotional aspect to the machine's expression, can have on the degree of interest of the participants towards the machine, will be studied. The project can be divided into separated blocs as shows in fig.1 .

## Environment features analysis

The experiment will take place in a delimited room. The room will be under a 3D analysis. Attention (object, gesture, posture and sound -oriented) algorithms will be used to detect the most salient areas and extract features from them. The features will be analyzed and classified using machine learning techniques to either one of the different scenarios defined (see section below) or to a “nothing happening” state. **Work package 1** will focus on that system of attention and will deliver exploitable features while **work package 2** will contain the creation of a suitable machine learning system capable of classifying with good enough precision the scenarios/state.

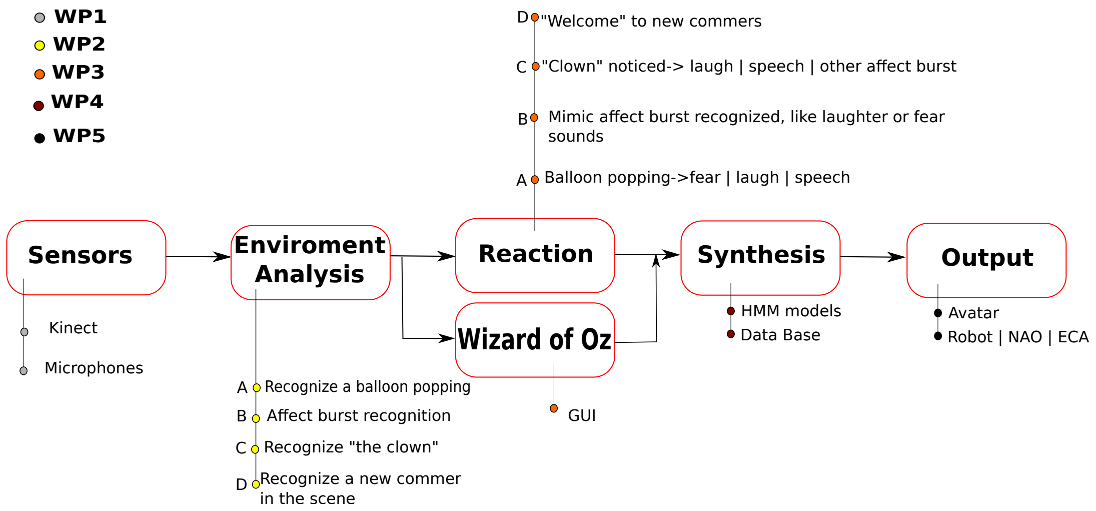

WP1
WP2
WP3
WP4
WP5
D "Welcome" to new commers
C "Clown" noticed-> laugh | speech | other affect burst
B Mimic affect burst recognized, like laughter or fear sounds
A Balloon popping->fear | laugh | speech
Sensors
Enviroment Analysis
Reaction
Synthesis
Output
Wizard of Oz
Kinect
Microphones
A Recognize a balloon popping
B Affect burst recognition
C Recognize "the clown"
D Recognize a new commer in the scene
GUI
HMM models
Data Base
Avatar
Robot | NAO | ECA

Figure 1: Pipeline of temporary general of the project

### Scenarios

In the framework of studying the avatar's reactions' impacts on the participants interests, several scenarios will be put in place. They are described as follows:

1. Welcoming a new participant coming in the room.
2. Expressing fear when sudden high enough sound occurs (e.g. balloon popping).
3. Mimicking "contagious" affect bursts (e.g. laughter or fear shouts).
4. Laugh and stare at a detected person appearing in a back door of the room and not noticed by the participants (this participant is referred to as "clown" in the diagram).

The most relevant scenario will be chosen directly from the decision made by the analysis part.There will also be a possibility to trigger some reactions using a wizard of Oz system.

## Avatar reaction

The avatar will react according to the scenario/state detected. Its response will be a combination of facial and vocal affect bursts expression. We expect the participants to also have a reaction which will trigger another reaction from the avatar, thus ending up with a few succession of reactive interactions between the avatar and the participants.

### Reactions

As mentioned previously, affect bursts will be displayed along with speech sentences by the avatar. They will be expressed as multimodal (vocal and corresponding facial expressions) affect bursts. The avatar's eyes direction will also be a type of reaction. The eyes will be oriented either towards the participants or in another direction. This will allow help to study the influence the avatar's reactions have on the participants interest and curiosity.**Work package 3** will take care of choosing the right reaction to synthesize and to create a wizard of Oz system in parallel of the automatic one.

## Synthesized reaction

A pre-recorded multimodal database will be used to synthesize the required reactions.The database contains different multimodal affect bursts: vocal affect burst recorded and their corresponding facial sensor coordinates. It also contains vocal speech recordings along with their corresponding facial sensor coordinates. The facial and vocal reactions will be synthesized either from HMM models created from the database or by extracting the desired sounds directly from the database. In both cases, the vocal synthesis will be synchronized with the facial expression one. **Work package 4** will focus on the synthesizer part, i.e. creating the HMM models, and make sure that the synthesis (of both vocal and facial expression parameters) are generated in real in time. **Work package 5** will deliver the avatar needed for the experiment, making sure that it can be driven in real-time by the parameters generated.

## Attention based Smart Room

A network of RGBD sensors will be used to create a 3D environment. From this large stream of 3D data, a multimodal analysis will be performed on images and pointclouds information to retrieve cues on users' attention, involvement, interest,etc. A rarity based computational attention mechanism specially developed for treating 3D data will be used as the core element of our smart room.

## References

[1] J. Urbain, H. Çakmak, and T. Dutoit, "Evaluation of HMM-based laughter synthesis," in *Acoustics Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.

[2] H. Çakmak, J. Urbain, J. Tilmanne, and T. Dutoit, "Evaluation of HMM-based visual laughter synthesis," in *Acoustics Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014.

[3] H. Çakmak, J. Urbain, and T. Dutoit, "Hmm-based synthesis of laughter facial expression," *Transactions on Affective Computing (TAC)*, 2015, [Submitted].

[4] K. Rapantzikos, G. Evangelopoulos, P. Maragos, and Y. Avrithis, "An audio-visual saliency model for movie summarization," in *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, Oct 2007, pp. 320–323.

[5] A. Zlatintsi, P. Maragos, A. Potamianos, and G. Evangelopoulos, "A saliency-based approach to audio event detection and summarization," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug 2012, pp. 1294–1298.

[6] Antoine Coutrot and Nathalie Guyader, "An audiovisual attention model for natural conversation scenes," in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 1100–1104.

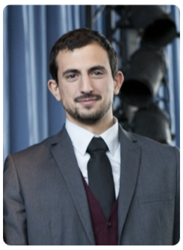
[7] Kai-Hsiang Lin, Xiaodan Zhuang, Camille Goudeseune, Sarah King, Mark Hasegawa-Johnson, and Thomas S. Huang, "Saliency-maximized audio visualization and efficient audio-visual browsing for faster-than-real-time human acoustic event detection," *ACM Trans. Appl. Percept.*, vol. 10, no. 4, pp. 26:1–26:16, Oct. 2013.

[8] Hyun S. Park, Eakta Jain, and Yaser Sheikh, "3d social saliency from head-mounted cameras," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds., pp. 422–430. Curran Associates, Inc., 2012.

[9] Raquel Viciana-Abad, Rebeca Marfil, Jose M Perez-Lorenzo, Juan P Bandera, Adrian Romero-Garces, and Pedro Reche-Lopez, "Audio-visual perception system for a humanoid robotic head," *Sensors*, vol. 14, no. 6, pp. 9522–9545, 2014.

[10] Tomoki Tsuchida and Garrison W Cottrell, "Auditory saliency using natural statistics," in *Proc. Annual Meeting of the Cognitive Science (CogSci)*, 2012, pp. 1048–1053.

[11] Kevin El Haddad, Stéphane Dupont, Jérôme Urbain, and Thierry Dutoit, "Speech-laughs: An HMM-based approach for amused speech synthesis," in *Acoustics Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, in press.



**Hüseyin Çakmak** holds a double degree in Aeronautics from the Higher Institute of Aeronautics and Space (ISAE) and in Electrical Engineering from the Polytechnic Faculty of Mons (FPMS). He is currently PhD student under a FRIA grant. His research interests are audio and visual synthesis and more specifically audiovisual laughter synthesis based on a statistical approach.

**Kévin El Haddad** holds a Master degree in microsystems and embedded systems from the Lebanese University. He is currently PhD student at the TCTS lab. of the Polytechnic Faculty of Mons (FPMS). He works on Affect Bursts analysis and synthesis in the framework of the European project JOKER.

**Nicolas Riche** holds an Electrical Engineering degree from the University of Mons, Engineering Faculty (since June 2010). His master thesis was performed at the University of Montreal (UdM) and dealt with automatic analysis of the articulatory parameters for the production of piano timbre. He obtained a FRIA grant for pursuing a PhD thesis about the implementation of a multimodal model of attention for real time applications.

**Julien Leroy** holds an Electrical Engineering degree from the University of Mons, Engineering Faculty. He is currently finishing his PhD in signal processing at the Circuit Theory and Signal Processing Lab (TCTS Lab/UMONS). His research interests include but are not limited to behavior analysis and 3D attention.