Nowadays people tend to communicate with computers as if they were real humans. In order to improve this interaction, researchers introduced autonomous on-screen characters capable of carrying on conversations with humans by both understanding and producing speech, facial expressions, gaze, hand gestures or body movements. They are called Embodied Conversational Agents (ECAs), a type of multimodal interface. This research branch brings promising future in many Human-Computer Interaction applications, such as in entertainment, games industry, information provider or guider, sales assistant, avatar in telepresence and virtual reality, healthy services, etc., but it also raises a lot of challenges in the construction of useful and communicative ECAs. Actually, it involves interdisciplinary research, from 3D computer graphic, affective computing to psychology, linguistics and anthropology. In this presentation, we would like to introduce our ongoing research project in developing an ECAs system within our laboratory at Telecom ParisTech.

For many years, the Multimedia Group at Telecom ParisTech has developed an ECAs system, namely GRETA that enables to produce and to respond appropriately verbal and nonverbal behaviors to the personality and emotional state of the users. Indeed, our ECAs system has been carried out based on SAIBA, (Situation, Agent, Intention, Behavior, Animation), an international standard multimodal behavior generation framework for ECAs.

Recently, the advance of humanoid robot technology allows them to have behavior capacities as much as the virtual agents do (e.g., the artificial face expressions of the Kismet robot at MIT, the dynamic body movements of ASIMO robot at Honda, human-like appearance of androids). Our ECAs system has been hence extended to be able to handle both virtual and physical agents.

In our ECAs, system takes as input a text to be said by the agent. The text has been enriched with information on the manner the text ought to be said (i.e. with which communicative acts the text should be transmitted). The system uses information extracted from the enriched text such as its structure, its semantic variety, certain pragmatic elements of the utterance as well as from emotional states and personality of agent's profile to select and compute corresponding non-verbal behaviors (e.g. hand gestures, facial expressions, etc.) to be realized by the agent. A speech synthesizer is integrated in the system to generate natural speech and return temporal information which allows the system to be setup the synchronization between multimodal behaviors.

Our presentation includes four sections. The first one gives a state of the art of ECAs and its potential applications. In the second one, we describe the general architecture of a SAIBA based ECAs system thus we explain how we have implemented on our Greta. We discuss about the remaining challenges of ECAs and our ambition in the third one. Finally, some videos of demonstration will help you have a clear picture of our work.