Proposition de recherche doctorale

Machine learning approach applied to multimodal behavior generation for virtual character

Mots clés :
● Directeur de thèse : Catherine PELACHAUD
● Co-encadrant(s) :
● Unité de recherche : Institut des Systèmes Intelligents et de Robotique
● Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
● Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)

Embodied Conversational Agents ECAs are virtual entity with human-like appearance. They also communicate verbally and nonverbally. They are used as interface in human-machine interaction taking several roles, such as assistant, tutor, or companion. They are endowed with communicative capability, that is, they can dialog with humans using verbal and nonverbal means. In this PhD we will focus on coverbal gestures that are gestures occurring during speech. These gestures are described along several parameters such as the movement of the hands (its path (planar, curved…), its dimension (X, Y, Z)), the hand shape and the wrist orientation (Calbris, 2011). During communication, facial expression, head movement, gestures participate in conveying meaning as much as speech. A pointing gesture indicates the object being discussed, a raise eyebrow emphasizes a word, a nod can mean agreement. Verbal and nonverbal behaviors come from a same planning process. They are tightly coupled, showing high synchronization mechanism. The speaker can indicate the shape of a box while talking about it. Doing such an iconic gesture may be more efficient than using solely verbal means to describe it. A common taxonomy used by scholars working on gestures defines 5 types of coverbal gestures: - Iconics that depict physical property of an object (eg its size) - Metaphorics that are similar to iconic but for abstract idea (eg a precision gesture) - Deitics that point to a direction, object, person - Beats that rhythm speech - Emblems that are highly lexicalized and conventional (eg the ‘ok’ gesture) So far most existing ECA behavior models have relied on creating a repertoire of nonverbal behaviors where each entry is a pair of a communicative act and its corresponding list of nonverbal behaviors. Several techniques have been deployed to create such a repertoire. Many of them rely on the analysis and annotation of video corpora. Others followed a user-centered approach where users are asked to create on the virtual agent the desired behaviors. Lately motion capture is used to gather precise body and facial motion. However, most of these existing techniques require defining ahead of time the shape of the behaviors and to which communicative acts they correspond to. Lately several machine learning (HMM, CRF…) have been applied to capture the link between prosody and beat gestures (Levine et al, 2010), prosody and upper body movement (Ding et al, 2013; Busso et al, 2005), pragmatics analysis and behaviors (Marsella et al, 2013). Chiu & Marsella (2014) developed two models; one that learns the mapping from speech to gesture annotation and the other that learns the mapping from gesture annotation to gesture motion. Lhommet & Marsella (2016) further looked in modeling gesture forms of metaphoric gestures using the image schema representation. These approaches gave interesting results, especially regarding the computation of the gesture timing. However they lack in capturing the link between speech content and gesture shape. In this PhD the aim is to develop further works relying on statistical approach. In particular it will focus on modeling coverbal gestures linked to speech acts, paying particular attention of capturing gesture shapes. The foreseen approach will rely on the analysis and annotation of an existing corpus in terms of speech act and gesture. Several steps are foreseen: 1) Get acquainted with the literature on gesture studies and ECA behavior models 2) Annotate existing corpus (NoXi database: https://noxi.aria-agent.eu) in terms of speech act, prosody feature and hand gesture. Whenever possible, we will rely on automatic annotation. In particular this can be applied for prosodic features using tools such as PRAAT, Prosogram… Speech act will be annotated using the ISO - DIT++ taxonomy (https://dit.uvt.nl/). Gesture shape will be defined using Calbris’s gesture feature representation (Calbris, 2011). 3) Develop machine learning that captures the link between prosody, speech act, gesture timing and gesture shape. The model will aim at determining core gesture shapes (eg the path of the hands, or their shape) that are associated to speech act. 4) Evaluation of the model will be done by replicating the computed coverbal gestures onto a virtual agent. We will use the Greta ECA platform (http://www.tsi.telecom-paristech.fr/mm/themes/equipe-greta/). We will ensure the gesture model follows the “ideational unit” properties as defined by Calbris’ theory. Perceptive and objective evaluation studies will be conducted.

Résumé du projet de recherche (Langue 2)

A particular challenge in creating ECAs is to compute automatically not only what to communicate but also how to do it, through which nonverbal behaviors. When conversing with a human user, ECA needs to perceive and understand what the user says. It then needs to plan how to answer him. If the user asks a question, it ought to respond to it; if the user gives an opinion, it can agree with it or not. Dialog models have been built to give to ECAs these dialogic capacities. However these dialog models focuses on the verbal content. They do not provide multimodal information, that is, which nonverbal behaviors accompany speech.

Informations complémentaires (Langue 1)
The behavior model developed in this PhD will be integrated within the Greta platform that is used in several national and European projects. The PhD student will collaborate with people in the lab as well as with collaborators of the ongoing projects (http://aria-agent.eu/).

Informations complémentaires (Langue 2)

- Stacy Marsella, Yuyu Xu, Margaux Lhomme, Andrew Feng, Stefan Scherer, and Ari Shapiro, "Virtual Character Performance From Speech", in Symposium on Computer Animation, July 2013.