Towards Enhancing Human-Robot Relationship: Customized Robot's Behavior to Human's Profile

Amir Aly and Adriana Tapus
Cognitive Robotics Laboratory - ENSTA ParisTech
828 Boulevard des Marechaux
91120 Palaiseau Cedex, France
{amir.aly, adriana.tapus}@ensta-paristech.fr—

Abstract

A social robot should be able to understand human's profile (i.e., human's emotions and personality), so as to make the robot able to behave appropriately to the multimodal interaction context. This research addresses the online recognition of emotions based on a new fuzzy-based methodology. It also focuses on investigating how could a match between the human's and the robot's personalities influence interaction. Furthermore, it studies the automatic generation of head-arm metaphoric gestures under different emotional states based on the prosodic cues of the interacting human. The conducted experiments have been validated with NAO robot from Aldebaran Robotics and ALICE robot from Hanson Robotics.

Introduction

The increasing need for a social intelligent robot nowadays requires an advanced level of artificial intelligence functions that can enhance the interactive capabilities and the decisive criteria of the robot. In this research, we focus on making the robot able to model and understand human's emotions and personality in order to adapt its generated multimodal behavior to human's profile, which helps in building up a long-term human-robot relationship.

Research Overview

An overview of this research is illustrated in Figure (1). In this model, the captured audio-visual data are used to define the profile of the interacting human and to setup for generating an appropriate customized multimodal action. The outputs of the behavior generation subprocesses are modeled directly on the robot, in the presence of the personality control factor that adapts the robot's behavior to human's personality.

Online Detection of Human's Emotions

Fuzzy Logic imitates human logic by using a descriptive and imprecise language in order to cope with ambiguous input data. Emotion could be considered as a part of these complex systems that do not have clearly defined borders between some of its classes. The purpose of this system is to build-up a fuzzy model that can recognize emotions and decide whether an expressed emotion is conform to a previously learned emotion clusters, or it constitutes a new cluster (not learned before) that requires a new verbal and/or nonverbal action to be synthesized (Aly and Tapus 2012).

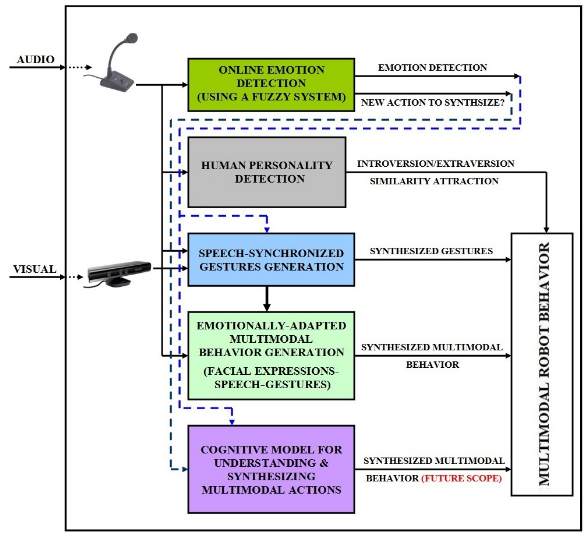

Figure 1: Research Overview

Customized Robot's Multimodal Behavior to Human's Personality

Personality plays an important role in the human-human communication and interaction. Various researches have demonstrated that personality is also crucial in human-robot interaction. In our recent work (Aly and Tapus 2013a), a direct relationship between the extraversion-introversion personality trait and the characteristics of the generated verbal and nonverbal behaviors has been found. The similarity attraction principle (i.e., individuals are more attracted by others who have similar personality traits) within a human-robot interaction context has been validated (Figure 2), by modeling a combined verbal and nonverbal robot behavior based on the extraversion-introversion personality dimen-

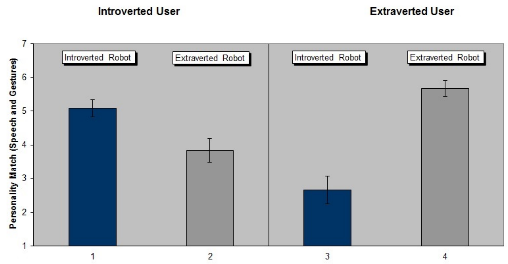

Figure 2: Personality Matching for the Introverted and Extraverted Robot's Conditions

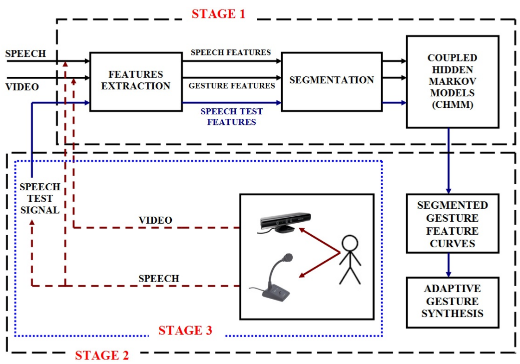

Figure 3: Speech-Synchronized Metaphoric Gestures Generation System

sion of the interacting human. The system's architecture contains different subsystems: (1) Personality recognizer toolkit, which estimates human's personality traits through performing a psycho-linguistic analysis on the dictated language, (2) PERSONAGE natural language generator, which adapts the generated text's content to human's personality dimensions, (3) BEAT toolkit, which translates the generated text into gestures (not including metaphoric gestures), (4) Metaphoric gestures generator, which uses prosody to generate synthesized metaphoric gestures, as illustrated in the next section (Aly and Tapus 2013b).

Customized Metaphoric Gestures Generation to Human's Emotions

The system is coordinated through 3 stages, as illustrated in Figure (3). Stage 1 represents the training stage of the system, in which the characteristics of the audio/video inputs get segmented and modeled on the CHMM, through which new adaptive head-arm metaphoric gestures are synthesized (i.e., stage 2) based on the prosodic patterns of a new speech-test signal which undergoes the same phases of the training stage. However, stage 3 represents an experimental stage in which the robot should be able to increase online its initial learning database by acquiring more raw audio and video data from humans in the surrounding of the robot (Aly and Tapus 2013b).

Figure 4: Synthesized Facial Expressions by ALICE Robot: Sadness, Happiness, and Anger (From Left to Right)

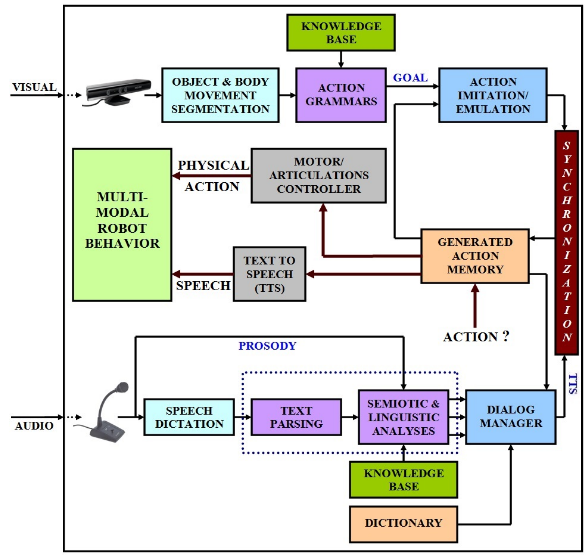

Figure 5: Computational model for understanding and generating multimodal actions

Emotionally-Adapted Multimodal Robot's Behavior Synthesis within a Storytelling

ALICE robot is used in the ongoing validation of the emotional expressivity of the generated gestures combined with facial expressions and expressive speech generated by MARY-TTS (text-to-speech toolkit) within an interactive human-robot storytelling (Figure 4), in which the generated multimodal robot's behavior should adapt the emotional context of the story.

Future Research Perspective

As referred to in Figure (1), future research orientations include developing a computational cognitive architecture that can simulate the functionalities of the Wernicke's area (Ojemann et al. 1989) and the Broca's area (Schaffler et al. 1993) in human brain which allow understanding and generating speech appropriately to the context of interaction, in addition to the mechanism of action imitation and emulation and the functionality of the mirror neurons that allow understanding and generating actions. A preliminary computational model is illustrated in Figure (5), in which the captured speech undergoes a semiotic and linguistic analysis in order to extract the semantic and pragmatic information, then the dialog manager phase which formulates a corresponding text to the context of interaction. On the other hand, action grammars (Pastra and Aloimonos 2012) are used to understand the goal of the captured actions, which will be learned by imitation or by emulating their goals. The processed synchronized speech and gestures will be stored in the multimodal action memory, ready for generating a future multimodal robot behavior whenever required.

Conclusion

The previously shortly discussed current research points aim towards creating an intelligent cognitive system that can make a robot more capable of understanding human's profile, so as to create a more natural, intuitive, and customized human-robot interaction relationship.

References

Aly, A., and Tapus, A. 2012. Towards an online real time fuzzy modeling for human internal states detection. In *Proceedings of the 11th International Conference on Control, Automation, Robotics, and Vision (ICARCV)*.

Aly, A., and Tapus, A. 2013a. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

Aly, A., and Tapus, A. 2013b. Prosody-based adaptive metaphoric head and arm gestures synthesis in human robot interaction. In *Proceedings of the 16th IEEE International Conference on Advanced Robotics (ICAR)*.

Ojemann, G.; Ojemann, J.; Lettich, E.; and Berger, M. 1989. Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Neurosurgery* 71:316–326.

Pastra, K., and Aloimonos, Y. 2012. The minimalist grammar of action. *Philosophical Transactions B* 367:103?117.

Schaffler, L.; Luders, H.; Dinner, D.; Lesser, R.; and Chelune, G. 1993. Comprehension deficits elicited by electrical stimulation of broca's area. *Brain* 116:695–715.