Toward a Universal Platform for Integrating Embodied Conversational Agent Components

*Hung-Hsuan Huang[1], Tsuyoshi Masuda[1], Aleksandra Cerekovic[2], Kateryna Tarasenko[1],
Igor S. Pandzic[2], Yukiko Nakano[3], and Toyoaki Nishida[1]
[1]Department of Intelligence Science and Technology, Graduate School of Informatics,Kyoto University, Japan
[2]Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[3]Department of Computer, Information and Communication Sciences, Tokyo University of Agriculture & Technology, Japan
*Presenter, E-mail: huang@ii.ist.i.kyoto-u.ac.jp

Demands of a Generic Platform for ECA Research

- Development involves multiple research disciplines → sophisticated and difficult for individual to develope
- No common interface standard → software components are not designed to cooperate with each other
- Ad hoc designs of integrating architectures → low reusability

Generic ECA Framework

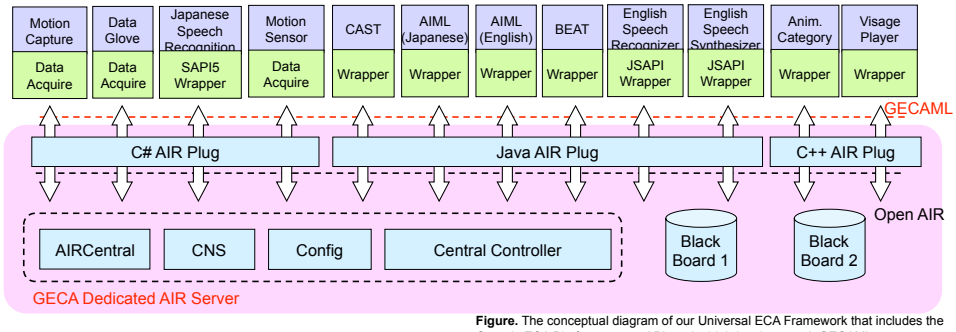


Figure. The conceptual diagram of our Universal ECA Framework that includes the Generic ECA Platform server, API, and a high-level protocol, GECAML

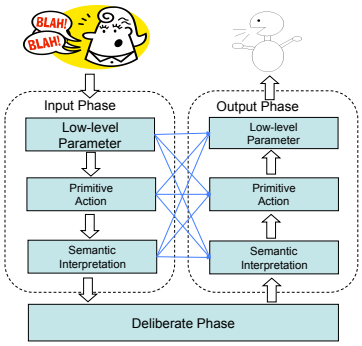


Figure. The hierarchy of the high-level protocol, UECAML, notes that reflex action links are shown in blue arrows

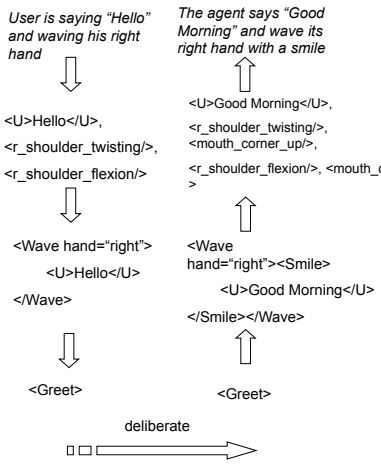


Figure. Example messages for a greeting response of the conversational to a human user's greeting behavior

Basic Concepts:

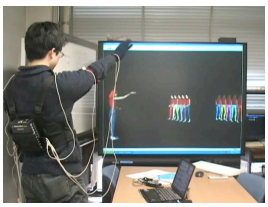
- A general purpose black-board based platform to mediate and transport messages among ECA software components
- A high-level protocol (GECAML) based on XML messages that is used in the communication between a standardized set of ECA components
- An application programming interface (API) for easily adapting ECA software modules to be plugged into the platform

Features:

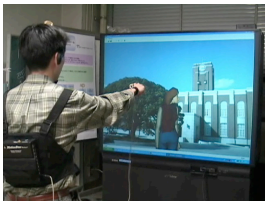
- Distributed architecture absorbs the differences of operating systems and programming languages of components and improves system performance
- Single-layer topology provides the possibility to support reflexive behaviors
- The weak inter-connectivity of the components allows the online switching of
- The Components and thus makes online upgrading and maintaining of components
- The Components have different levels of complexity and function can be directly integrated into the ECA system if they communicate in the same message types

Experimental Systems

(a)

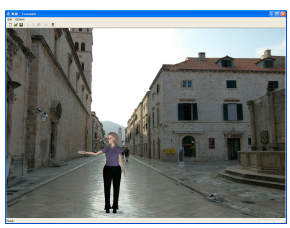


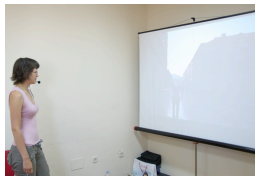
(b)

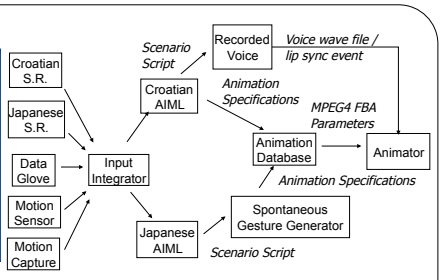


(a) The application for experiencing cross-culture gesture difference is a virtual environment contains one user avatar and multiple embodied agents. The avatar reproduces the user's hand gestures such as beckoning while the embodied agents react to those gestures pretending they come from different countries, Japanese or British. The user's actions are captured by a magnetic motion capturing device and translated to low-level joint angle parameters to drive the avatar character in real-time. The ten embodied agents are droved by ten reflexive controlling units individually with a common BAP catalogue component and ten individual figures those are driven by low-level MPEG-4 BAPs sent from the blackboard

(b) In this campus guide system, there is an embodied agent who stands in front of a background image; say a photo of the entrance of a university. The human user can ask the agent to explain what an object is in the background image with speech, pointing a location on the screen with right hand and head movements. The pointing gesture and its position is detected by the result from a magnetic motion capturing device and the head movements is detected by a 3-axis acceleration sensor







The screen capture, component configuration, and a scene of user-agent interaction of the Dubrovnik tour guide agent application. This application is developed by a student teamwork project titled as "An Agent Based Multicultural User Interface in a Customer Service Application" during the four-week international summer workshop, eNTERFACE'06 that was held in Dubrovnik, Croatia, 2006. In the planned system, the agent switches between its Croatian mode and Japanese mode by the speech recognition result. It speaks and performs cultural specific gestures in each mode. Because of the lack of Croatian speech recognizer and synthesizer, the speech input / output is done by English speech recognizer and prerecorded voice tracks. Due to the limited time, it couldn't be completed and the final version works in English and Japanese mode