Towards Evolutionary Multimodal Interaction

Maria Chiara Caschera, Arianna D'Ulizia, Fernando Ferri, and Patrizia Grifoni

Institute of Research on Population and Social Policies (IRPPS) –
National Research Council (CNR)
00185, Rome (Italy)
{mc.caschera,arianna.dulizia,fernando.ferri,
patrizia.grifoni}@irpps.cnr.it

Abstract. One of the main challenges of Human Computer Interaction researches is to improve naturalness of the user's interaction process. Currently two widely investigated directions are the adaptivity and the multimodality of interaction. Starting from the adaptivity concept, the paper provides an analysis of methods that make multimodal interaction adaptive respect to the final users and evolutionary over time. A comparative analysis between the concepts of adaptivity and evolution, given in literature, is provided, highlighting their similarities and differences and an original definition of evolutionary multimodal interaction is provided. Moreover, artificial intelligence techniques, quantum computing concepts and evolutionary computation applied to multimodal interaction are discussed.

Keywords: Multimodal interaction, Evolutionary computation, Adaptivity.

1 Introduction

In the next years, human-computer interaction will go towards a flexible and natural human-like interaction. Multimodality allows flexible interaction between users and systems because users have the freedom of using the modality of interaction of their choice (such as speech, handwriting, gestures, and gaze), and they have a sense of naturalness that derives from the use of interaction modes similar to the ones used in everyday human-human interactions.

One main challenge in multimodal interaction design lies in the adaptation of the interaction process to the users. In detail, the challenge is in selecting the optimal set of modal inputs combination that users will find easy and intuitive to produce, and that the system will be able to interpret in order to improve the degrees of flexibility and naturalness. The concept of adaptivity implies to decide on the appropriate devices, the appropriate interaction modalities according to user's profile information. User profiles include information about: user's preferences, rules, and settings (like name, address, birth date, or credit-card number); device and service profiles (like audio volume and text display properties); situation-dependent profiles (such as the context types such as environment and relevant changes within the current situation) [1]. User profiles contain dynamic information due to the fact that properties can be added, refined or deleted either manually/explicitly by the user or automatically/ implicitly by the system analysing the user's interaction process. This implies that,

P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, pp. 608–616, 2012.

during the interaction process, it is not sufficient to consider only issues related to adaptation but also evolutionary aspects. In fact, the user's profile evolves, and its evolution is due to the improvement, refinement and adaptation to changing of the user's preferences.

Although the naturalness and flexibility of human communication is hard to achieve, interactive systems will aim to these two main characteristics in a threefold way: through multimodality, adaptation, and evolution.

The naturalness and flexibility of multimodal interaction is motivated by the fact that humans communicate multimodally with their five senses. Human brain, indeed, processes multiple streams of information to assess the state of the world. The synergistic processing of these streams results in better performance and less cognitive load for the user compared to a situation in which s/he uses a single modality.

Analogously, the naturalness and flexibility of adaptive interaction is motivated by the fact that it supports heterogeneous user groups with variable and different needs, abilities and preferences by representing, reasoning, and acting on models of the user, domain, task, discourse, and media (e.g. graphics, natural language, gesture) [2].

Finally, the naturalness and flexibility of evolutionary interaction is motivated by the fact that human interaction is evolutionary by its nature, since all signs used during dialogues (language, gestures, eye movements, etc.) change and evolve over time.

Starting from the concepts of multimodality, adaptation, and evolution, the paper provides an analysis of methods applied for multimodal, adaptive and evolutionary interaction. In particular, the proposed work aims to investigate how artificial intelligence techniques, and in particular evolutionary computation, can be applied to make multimodal interaction evolutionary.

The remainder of the paper is structured as follows. The state of the art on multimodal interaction is presented in Section 2. Section 3 provides some notions and literature approaches on adaptive interaction. A comparative analysis between adaptivity and evolution is provided in Section 4, along with evolutionary algorithms applied in the human-computer interaction literature. Section 5 offers a discussion about research goals towards an evolutionary multimodal interaction. Finally, conclusions are presented in Section 6.

2 Multimodal Interaction

Multimodal interaction has emerged as the future paradigm of human computer interaction (HCI). This fact is gathered also by the increasing application of the multimodal paradigm to computer interfaces in order to make computer behaviour closer to human communication.

Communication among people is often multimodal as it is obtained combining different modalities, such as speech, gesture, facial expression, sketch, and so on. Multimodal interfaces allow several modalities of communication (see Fig.1) to be harmoniously integrated, making the system communication characteristics more and more similar to the human communication approach. The main features of multimodal interaction, such as the different classes of cooperation between different modes, the time relationships among the involved modalities and the relationships between chunks of information connected with these modalities, are described in [3].

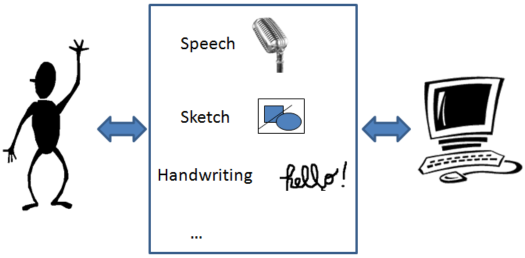

Fig. 1. Multimodal interaction

One of the main problems fundamental to the design of multimodal interfaces is the multimodal data fusion, i.e. the process of combining information from different modalities in order to have a comprehensive representation of the user's message. In the literature, three main different approaches to the fusion process have been proposed, according to the main architectural levels (recognition and decision) at which the fusion of the input signals can be performed: recognition-based (named also early or feature-based fusion), decision-based (named also late or semantic fusion), and hybrid multi-level fusion [4]. A more extensive discussion about multimodal input fusion strategies can be found in [5].

Analogously to the fusion process, another critical issue in multimodal interaction systems is the multimodal fission, i.e. the process of combining different outputs from modal channels in order to provide the user with consistent feedback. Foster [6] defines fission as *"the process of realising an abstract message through output on some combination of the available channels"*. A discussion about multimodal fission approaches can be found in [7].

The interpretation process is one important unit for building multimodal systems. The interpretation of user input is strictly connected to different features, such as available interaction modalities, conversation focus, and context of interaction. A correct interpretation can be reached by simultaneously considering semantic, temporal and contextual constraints. For example, in multimodal system based on video and audio input [8], the interpretation defines a multimodal corpus of digital and temporally synchronized video and audio recordings of human monologues and dialogues. An overview of methods for interpreting multimodal input is provided in [9].

3 Adaptive Interaction

Interaction techniques are defined as methods that allow a user to accomplish a given task via the user interface [10]. During the interaction process, every user has a certain physical ability level. An adaptation to the user's specific attributes implies to change certain parameters of the user interfaces and to adjusting its sensitivity according to the user's need. To adapt to user diversity, during interaction process, user models becomes crucial to provide adaptation and personalization for users.

In [11] three types of adaptation are presented: switching between interaction techniques; enhancing the interaction technique with modalities; and adapting the interaction technique itself.

During the interaction process, users may encounter different situations (e.g. environment condition, position of the user) that might influence the performance of executing the task. Switching between interaction techniques enables user to perform a task fitting the different environmental conditions.

The interaction process can be adapted by enhancing it with the use of multimodal feedback such as visual, audio or haptic [12].

Finally, adapting the interaction technique itself implies constructing, maintain and exploiting explicit representations of user. These models contain information and assumptions about users that system believes to be true, such as users' interaction patterns, preferences, interests, goals and characteristics. In detail, a user model combines the information provided by the user, direct inferences from the user's actions, and predictions made by stereotypes or user group models that are believed to be appropriate for this user [13].

Information for building the user model can be obtained simply asking the user for her/his preferences, or automatically inferring user habits, by learning user profile data. An example of automatic inference of user behaviours is proposed in [14] where users' knowledge and goals are used to automatically discover user profiles by analysing the tags users associate with available content. Moreover, contextual aspects, such as location and device constraints, are used in [13] where context-aware user profiles are proposed, in which the profile definitions are associated with particular situations encountered by the users.

Methods, those directly ask the user for her/his preferences, allows simply gather information for a user model. However, those methods are not always a trustworthy source of information about users because people tend to give socially acceptable or desirable answers or they may not be able to answer.

On the other hand, methods, those automatically infer user habits, allow constructing user models by making direct inferences from a user's behaviour to a model of the user. Disadvantages of those methods are: the difficulties to make inferences based on the available input; scarcity of data that is available; the difficulty to monitor user interactions; and the short-term knowledge that becomes outdated [12].

The limits of adaptive systems depend on some systems are targeted on user, others on task, others on system or environment and so these dependencies limit their extensibility and reusability [15].

4 From Adaptivity to Evolutionary

The concept of adaptive interaction, introduced in Section 3, is fundamental for designing interactive systems that are able to dynamically adapt the interaction to the user and context's features. This adaptation occurs for short-term changes, as the system reacts to situational changes, such as, for example, in context-aware systems.

However, the needs and conditions in which the interaction occurs can change over time due to various factors (for example, environmental or temporal). Therefore, this requires that the interactive system adapts and evolves the interaction process to these long-term changing circumstances.

In the literature the concept of "interaction evolution" has been introduced for representing the ability of interactive systems to evolve and tailor interaction in long-term changing situations. Interaction evolution is defined by McBryan et al. [16] as "multiple related instances of interaction configuration (customisation or personalisation) over time that have a goal to change some aspect of the systems interaction behavior". For interaction configuration the authors mean a combination of devices, interaction techniques, modalities used and supporting components required to instantiate a new configuration of the system.

In order to better specify the concepts of adaptive interaction and evolutionary interaction this work aims to make evident differences in terms of responses that the systems returns when unexpected stimuli are provided. For this purpose these definitions are provided.

Definition 1. *Adaptive Interaction* is characterized by an adaptive behavior of the system, i.e., the system reacts without producing a modification of its default functioning when it receives unexpected stimuli, but producing specific responses according to the different users or situations. ■

Definition 2. *Evolutionary Interaction* is characterized by an evolutionary behavior of the system, i.e. the system changes the default behavior when the same kinds of unexpected stimuli are frequent and repeated for a long time period, so that they become expected independently from the single user or situations features. ■

Fig. 2 shows the differences between adaptive (Fig. 2.a) and evolutionary interaction (Fig. 2.b). In the adaptive interaction, the system reacts in a twofold way, according to the kind of stimuli: for unexpected stimuli, it reacts with specific responses according to the user's features; for standard stimuli, it reacts with its default behavior (never changing). In the evolutionary interaction, the system reacts both to standard and unexpected stimuli with a default behavior (ever changing).

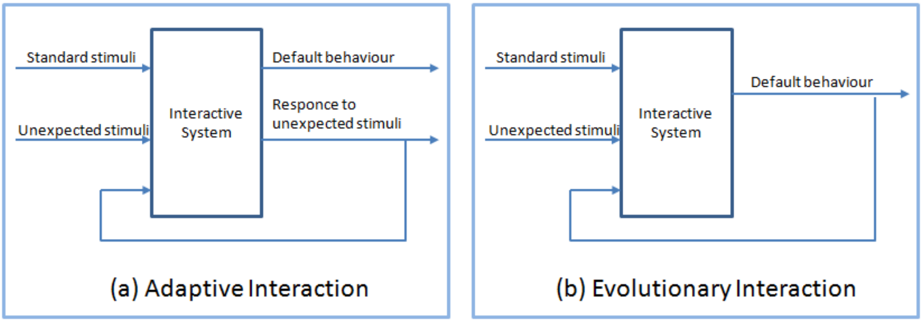

Fig. 2. Adaptive (a) and Evolutionary (b) Interaction

Interaction evolution is the result of the efforts made to endow interactive systems with a flexible human-like interaction. Human interaction, indeed, is evolutionary by its nature, since all interaction abilities and features involved in dialogues (language, gestures, eye movements, etc.) change and evolve over time generation by generation.

Therefore, in recent years many works have focused on assuring flexibility and naturalness to human-system communication through the exploitation of the interaction evolution. In particular, McBryan et al. [16] proposed a model for representing the interaction evolution process, which is composed of one or more potentially linked interaction configurations, each of which consists of four sequential stages: identify opportunity, reflect, decide and implement. Goldin and Keil [17] provided some considerations about the relationships among interaction, evolution and intelligence. In particular, the authors stated that the paradigm shift from algorithmic computation towards evolutionary and interaction computing can contribute to establish the unified foundations of interactive systems, overcoming the limits of the algorithmic approaches in terms of intelligence.

Evolutionary computation is a field closely related to adaptation and evolution. Evolutionary algorithms, indeed, were proposed initially by Holland [18] to drive the adaptation of complex systems to changing and uncertain environments. They use nature inspired concepts, like mutation, recombination, and selection, applied to a population of individuals containing candidate solutions in order to evolve iteratively better and better solutions.

Therefore, evolutionary algorithms have been widely used for addressing adaptivity respect to users and contexts uncertainness and variations, and for optimizing adaptivity processes. According to the concept “evolutionary interaction”, previously introduced, the different algorithms, proposed in the literature, can act on two levels, both at the level of adaptability and, at the level of default operation of the interaction producing its evolution on the base of unexpected configurations. The reference population and the evolution are then considered respect to the default interaction. This perspective considers communication and interaction between humans changing with the influence of context (e.g. cultural and socio-economic) and individual.

Hence, evolutionary algorithms have been successfully applied in the human-computer interaction literature to facilitate user personalisation without the need for time consuming explicit knowledge-acquisition process. Pauplin et al. [19] proposed an interactive evolution approach for developing rapidly reconfigurable systems in which the users' tacit knowledge and requirements can be elicited and used for finding the appropriate parameters to achieve the required image segmentation. Toney [20] applied an evolutionary-based reinforcement learning algorithm for automatically generating spoken strategies in a spoken dialogue system. Another evolutionary-based learning classifier algorithm was proposed in [21], which uses information from a user’s environment for automatically personalizing desktop applications.

From the analysis of these evolutionary methods it is clear that they have been applied to specific domains for enabling a short-term adaptation to situational changes. However, considering the evolutionary nature of these methods, it is interesting to investigate the opportunity to apply them for modeling the interaction evolution in long-term changing situations.

5 Evolutionary Multimodal Interaction

Technology is changing, people are changing, society is changing, and HCI has to intervene and adapt itself to these changings. HCI needs to extend its methods and approaches in order to fit on human changing.

Considering human interaction, it changes and evolves through repeated use in order to adapt the features of interaction between people with differences in communication. The repetition of communication and the re-adaptation lead to evolution of the interaction process. The evolutionary nature of the human interaction has been investigated in [22] analyzing the repetition of miscommunication and the successful communication.

As well as human communication evolves, so the interaction between human and system evolves. Similarly to the natural world where genetic algorithms are used to analyze the evolution of species, so the evolutionary computation may be applied to multimodal interaction in order to observe the evolution of interaction.

As proof of this consideration, similarly to humans, the interaction between robots has been investigated and modeled in [23] using Recurrent Neural Networks (RNN) in order to obtain a mutual adaptation between agent robots. Moreover, adaptive statistical methods have been applied in [24] to optimize the combination of multimodal user input.

Researches on evolutionary computation integrated with multimodal and adaptive interaction will play a key role for designing interactive systems that enable a flexible natural human-like interaction. These researches will improve the current concepts, theories and models of evolutionary interaction given in the literature (e.g. [23], [24]). Starting from the Evolutionary Interaction concept provided in Definition 2, an effort in this direction will allow: (i) to define new interaction models and algorithms able to represent and manage the evolution over time of interaction processes, its interpretation; (ii) to define evaluation metrics for measuring interaction naturalness and the accuracy of models representing the evolution in long-term changing situations.

Considering new methodologies for the design of evolution of interaction models, different evolutionary algorithms and computational approaches can be used.

We believe that a view of particular interest may be genetic algorithms evolved in the perspective of quantum computing.

Quantum Genetic Algorithm (QGA) uses concepts of qubit, quantum superposition, quantum logic gate and other concepts [25] [26]. Qubit is the unit of quantum information; it is a two state system and it exists in all its potential states, but when it measured gives one only of the possible configurations (superposition).

In classical computing, the possible states of a system of n bits form a vector space of n dimensions while, in a quantum system of n qubits the state space has 2^n dimensions. An interesting example of QGA for multisensors image registration has been proposed in [27], but their use could be extended to the interaction and communication processes, and in particular to the multimodal interaction. In fact, these algorithms allow to represent the complexity of communication processes and, in particular, of multimodal input, even using populations having small size respect to the population for co-respective in traditional computing, providing an optimization in terms of computing time, and providing transformation and evolutions considering in a very efficient way and in parallel all coevolving communication issues.

6 Conclusion

This paper has discussed the roles of multimodality, adaptation and evolution in making interactive systems natural and flexible as human-human interaction.

In detail, the main features of multimodal interaction have been presented in order to underline how the naturalness can be achieved in HCI, emphasizing the importance of the adaptivity issue.

The paper presents the concept of adaptivity underlining how adaptation occurs for short-term changes. It underlines the necessity for interactive systems to consider the human changing in long time period and how they can be reflected in the interaction processes.

For this reason, the paper focuses on the evolutionary nature that the HCI has to consider in order to fit the changing and the uncertainties of human interaction.

Starting from the investigation on the application of evolutionary algorithms to users and contexts uncertainness and variations in short term adaptation, this paper addresses of using genetic algorithms evolved in the perspective of quantum computing to provide an evolutionary HCI.

References

1. Thomsen, J., Vanrompay, Y., Berbers, Y.: Evolution of context-aware user profiles. In: ICUMT 2009, vol. 1(6) (2009)
2. Maybury, M., Wahlster, W.: Intelligent User Interfaces: An Introduction. In: Readings in Intelligent User Interfaces, pp. 1–13. Morgan Kaufmann, San Francisco (1998)
3. Caschera, M.C., Ferri, F., Grifoni, P.: Multimodal interaction systems: information and time features. International Journal of Web and Grid Services (IJWGS) 3(1), 82–99 (2007)
4. D'Ulizia, A., Ferri, F., Grifoni, P.: Generating Multimodal Grammars for Multimodal Dialogue Processing. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40(6), 1130–1145 (2010)
5. D'Ulizia, A.: Exploring Multimodal Input Fusion Strategies. In: The Handbook of Research on Multimodal Human Computer Interaction and Pervasive Services: Evolutionary Techniques for Improving Accessibility, pp. 34–57. IGI Publishing (2009)
6. Foster, M. E.: State of the art review: multimodal fission. Public deliverable 6.1, COMIC project (2002)
7. Grifoni, P.: Multimodal fission. In: The Handbook of Research on Multimodal Human Computer Interaction and Pervasive Services: Evolutionary Techniques for Improving Accessibility, pp. 103–120. IGI Publishing (2009)
8. Harper, M.P., Shriberg, E.: Multimodal model integration for sentence unit detection. In: ICMI 2004, pp. 121–128 (2004)
9. Caschera, M.C.: Interpretation methods and ambiguity management in multimodal systems. In: Grifoni, P. (ed.) Handbook of Research on Multimodal Human Computer Interaction and Pervasive Services: Evolutionary Techniques for Improving Accessibility, pp. 87–102. IGI Global, USA (2009)
10. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces. In: Theory and Practice. Addison-Wesley (2005)
11. Renny Octavia, J., Raymaekers, C., Coninx, K.: Adaptation in virtual environments: conceptual framework and user models. Multimedia Tools Appl. 54(1), 121–142 (2011)

12. Ehlert P.A.M.: Intelligent user interfaces: introduction and survey. Research Report DKS03-01 / ICE 01 Data and Knowledge Systems group Faculty of Information Technology and Systems Delft University of Technology (2003)
13. Au Yeung, C.M., Gibbins, N., Shadbolt, N.: A Study of User Profile Generation from Folksonomies. In: Social Web and Knowledge Management, Social Web 2008 Workshop at WWW 2008, Beijing, China, April 21-25 (2008)
14. Carrillo-Ramos, A., Villanova-Oliver, M., Gensel, J., Martin, H.: Contextual User Profile for Adapting Information in Nomadic Environments. In: Weske, M., Hacid, M.-S., Godart, C. (eds.) WISE Workshops 2007. LNCS, vol. 4832, pp. 337–349. Springer, Heidelberg (2007)
15. Rousseau, C., Bellik, Y., Vernier, F.: Architecture Framework for Output Multimodal Systems Design. In: Proceedings of OZCHI 2004, November 22-24. ACM Press (2004)
16. McBryan, T., McGee-Lennon, M.R., Gray, P.: An integrated approach to supporting interaction evolution in home care systems. In: 1st International Conference on Pervasive Technologies Related to Assistive Environments, Athens, Greece, July 16-18, vol. 282, p. 1. ACM Press (2008)
17. Goldin, D., Keil, D.: Interaction, Evolution, and Intelligence. In: Proc. CEC 2001 (Congress on Evolutionary Computation), Seoul, Korea, vol. 2, pp. 805–814. IEEE Press, Seoul (2001)
18. Holland, J.H.: Adaptation in Natural and Artificial Systems. The Univ. of Michigan Press (1975)
19. Pauplin, O., Caleb-Solly, P., Smith, J.: User-centric image segmentation using an interactive parameter adaptation tool. Pattern Recognition 43(2), 519–529 (2010)
20. Toney, D.: Evolutionary Reinforcement Learning of Spoken Dialogue Strategies. PhD thesis (2007), `http://hdl.handle.net/1842/1769` (retrieved)
21. Shankar, A., Louis, S.J.: XCS for Personalizing Desktop Interfaces. IEEE Transactions on Evolutionary Computation 14, 547–560 (2010)
22. Miwa, Y., Wesugi, S., Ishibiki, C., Itai, S.: Embodied interface for emergence and co-share of 'Ba', usability evaluation and interface design. In: Proceedings of HCI Int., pp. 248–252 (2001)
23. Hinoshita, W., Ogata, T., Kozima, H., Kanda, H., Takahashi, T., Okuno, H.G.: Emergence of Evolutionary Interaction with Voice and Motion between Two Robots using RNN. In: IROS 2009 Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4186–4192 (2009)
24. Althoff, F., Al-Hames, M., McGlaun, G., Lang, M.: Towards a New Approach for Integrating Multimodal User Input Based on Evolutionary Computation. In: Proceedings ICASSP 2002. IEEE Signal Processing Society, Orlando, USA (2002)
25. Meshoul, S., Batouche, M., Belhadj-moustefa, K.: An evolutionary framework for image data fusion based on the maximization of mutual information. In: Proceeding of the International Symposium on Software and Systems, I3S 2001 (2001)
26. Han, K., Kim, J.: Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE Transactions on Evolutionary Computation 6(6) (2002)
27. Draa, A., Batouche, M., Talbi, H.: A quantum-inspired differential evolution algorithm for rigid image registration. Trans. Eng. Comput. Technol., 408–411 (2004)