Multimodal Interfaces: An Enactive Approach

 Beatriz A. Pacheco  
 Mackenzie Presbyterian University  
 Computer and Informatics Faculty  
 Rua da Consolação, 930 CEP 01302-907  
 São Paulo – SP - Brazil  
 bia.pacheco@mackenzie.br

 Ilana A. Souza-Concilio  
 Mackenzie Presbyterian University  
 Computer and Informatics Faculty  
 Rua da Consolação, 930 CEP 01302-907  
 São Paulo – SP - Brazil  
 iasouza@mackenzie.br

Abstract— Cognitive infocommunications (CogInfoCom) study the connections between infocommunications and cognitive experiences. This paper matches with the purpose of the Conference because it explores Enactive knowledge and its relations with Interaction Design of Multimodal Interfaces. Enactive knowledge is what it is built through the action and is built on motor skills as manipulating an object, riding a bike or practicing sports. It is acquired by doing. On other hand, multimodal interfaces are known as inherently flexible, and to provide an especially ideal interface for accommodating both the changing demands on technology and the large individual differences present in the population. It’s possible to affirm that the main concept of enactive interfaces is to allow the construction of knowledge and storage of information from interfaces that enable some kind of motor action. This paper aims to discuss the exploratory process of the relations between men and multimodal systems, prospecting new possibilities from an enactive vision and analyzing how it can substantially reduce the gap still existing between men and technology, especially in their everyday objects and devices.

Keywords— Cognition; Multimodal Interfaces; Interaction; Enaction.

I. INTRODUCTION

Cognitive Infocommunications (CogInfoCom) study the relations between cognitive experiences and communication mediated by technology. From the junction of this two research areas, several of innovative application approaches have been emerging. According to Csapó and Baranyi [1], “the primary aim of CogInfoCom is to better understand how natural cognitive processes can effectively co-envolve with processes in artificially cognitive systems”.

Accordingly to such engagement, a possible approach that may add value to CogInfoCom research is the Enaction theory proposed by Maturana and Varela [2].

For Varela, Thompson and Rosch [3], there are three moments to be considered in the exploration of cognitive Science: the cognitivism, the connectionism and the enaction. The main idea of cognitivism is that the intelligence is so similar to computing in its main characteristics that cognition can be defined by computations about symbolic representations. In connectionism, the brain is considered to operate in a distributive way based in massive interconnections, in a way that the effective connections between sets of neurons are modified due to the progress of experience.

Finally, enaction, term coined by the Chilean biologists Humberto Maturana and Francisco Varela [2], from the Spanish expression en acción that can be understood through two point of views: (1) perception consists in an action guided by perception, or the research of perception is the study of how the perceiving subject can guide its actions in a local situation and (2) the cognitive structures emerge from recurrent sensorimotor schemes that allow the action to be guided by perception. The sensorimotor structure is “the manner in which the perceiving subject is inscribed in a body, [...] that determines how the subject can act and be modulated by the events of the environment.” [3].

In this sense, cognition is situated in the interaction of body and world, dynamic bodily process such as motor activity can be part of the reasoning process, and offline cognition is body-based too. Finally embodiment assumes that cognition evolved for action, and because of this, perception and action are not separate systems, but are inextricably linked to each other and to cognition. This last idea is close to the core idea of enaction [4].

Pacheco [5] affirms that enactive knowledge is what is built through the action and is built on motor skills as manipulating an object, riding a bike or practicing sports. Enactive knowledge is acquired by doing.

Thus, the foundations of multimodal interfaces lay in the facility of human users with the simultaneous diverse interactive streams. This facility, in turn, has its root in the multifacettens of human embodied sensing, perception, cognition and action [6].

As a research subject, multimodal interaction encompasses extrem spectrum of research domains, from cognitive psychology to software engineering, including human-computer interaction, which is already cross-disciplinary [7]. While cognitive psychologists study how the human brain processes information and interacts through various modalities, interaction practitioners are interested in how humans use multimodal interfaces, and software engineers are interested in building tools and systems supporting the development of such multimodal interfaces, thus studying software architectures and multimodal processing techniques.

According to Oviatt [8], multimodal interfaces target a more “human” way of interacting with computers, by means
of speech, gestures or other modalities, as well as being preferred over unimodal interfaces by users; multimodal interfaces also have been demonstrated to offer better flexibility and reliability than other human/machine interaction means.

Thus, this paper aims to discuss the exploratory process of the relations between men and multimodal systems, prospecting new possibilities from an enactive vision and analyzing how it can substantially reduce the gap still existing between man and technology, especially in their everyday objects and devices.

II. COGNITION BY ENACTION

Varela [9] use the term action to highlight that “the sensorial and motor processes, perception and action are fundamentally inseparable in lived cognition. It must be emphasized that they are not associated with individuals by simple contingency: the evolved together”. Thus, it can be stated that in this context, cognition is fundamentally a characteristic of living organisms in an adaptive and dynamics with the its environment.

In the enactive approach, cognition is no longer understood as a formal computation of symbols, or is considered as an appropriate solution to a given task, but the action of the subject is thought of as member of a complex network formed at multiple levels of interconnected subnetworks that overcomes the dichotomy between individual / internal versus environment / external: the cognitive system of the subject will be part of a durable yet changeable existing world, in which individual and environment are instances that co-imply.

Only recently the term enaction gained widespread acceptance in areas such as human-computer interaction. “Enactive Interaction between human and world (including humans and technologies) is seen as a process participating to [10]:

- the co-construction of the mind with the body
- the co-construction of the humans and the world. Mediated Artificial systems, called “Enactive Interfaces”, which preserve this type of interaction, would be favorable conditions
- to understand such complex processes
- to allow humans to produce and create in really fruitful way”.

In this sense, when thinking about development of enactive interface, we are facing a new kind of Human-computer Interfaces that allow expressing and providing enactive knowledge by means of the integration of different sensorial aspects. According to Hutchins [4], enaction is the idea to complement that our experience of the world is created in our organism, modeled by our actions. Thus, Varela et al. [3] affirm that is possible to identify five linked ideas that make the notion of enaction, that are autonomy, sense making, emergence, embodiment and experience.

What seems to be attractive in this perspective is to consider what might emerge from this design in the dialogue construction with technology, especially in a context of ubiquity, in which the compute is no longer a device restricted to desks. With the technology evolution, every object may become a potential computer, as it has the ability to manipulate and execute instructions.

Much of the technological ecology, embodied in the new digital artifacts suffered rough transformations in the last few years. The diversity, the omnipresence and the sizes of the screens began to change significantly, and the new ubiquity scenario (mobility + pervasiveness) is providing even greater enhancement of such artifacts, that are incorporating in their interaction paradigms (more natural) the gestures, movements, voices and sounds. But for that, we need to establish a new critical base to reshape the cognitive science into this new perspective, based on Embodiment and Enaction. These movements are already happening gradually in cognitive Science and in a more subtle way in computing, but it is required solidification so that designers can design in light of the new knowledge.

A relatively new research stream has been trying to build interfaces that work according to the enaction theory proposed by Varela et al [3]: interactive systems that allow organizing and transmitting this specific type of knowledge. In this sense, the multimodal interfaces are candidates to the creation of enactive interfaces due to the coordinated use of sound, touch and vision.

The key-element of enactive approach is that the sensorial knowledge and skills are necessary parts of perception [11]. This confirms the validity of the assertion of Merleau-Ponty that every theory of the body already is a perception theory. However, the explanations about such concept suffered severe transformations as time changed. Thus, it has become plausible to assure that perception is itself a mode of action. Noê [11] says that perception is not something that happens to us or in us; it is something that people do. What is perceived is determined by what is done; by what you know how to do or what you are capable of doing. These actions are subtly different but strongly related.

Perceiving, according to the author [11], is implicitly testing the effects of movement in sensorial stimulation. His most relevant statement is that there is an enactive action that is the ability to perceive itself, and this is not only dependent but built by the fact that people have a certain kind of sensorimotor knowledge.
A. Enactive interactions

The main concept of enactive interfaces is to allow the construction of knowledge and storage of information from interfaces that enable some kind of motor action. Thus, this kind of interfaces should be able to understand and transmit the user gestures to give an appropriate feedback in terms of perception. According to Enactive Network [12], such components can be considered a new step in the development of human-computer interaction because they are characterized by a closed circuit that understand the user natural gestures, that are that is the real challenge of current gestural interfaces, and the activated perceptual modalities. Those interfaces are conceived to explore the direct loop beyond the ability to recognize complex gestures.

The development of those interfaces require the creation of a common vision between different research areas, such as the development of haptic devices, computer vision and sound processing, providing more attention to the motor aspect of the interactive actions.

The enactive perspective provides a new structure for the understanding of the use of technology by humans, based in the specific comprehension of their biological foundations. It leads to new kind of design posture, the design of experience, which goal is no longer design human-computer interfaces that are the objects of perceptual interaction, but interfaces that suit as implicit (transparent) means that allow their users to give place to new ways to experiment the world.

Enactive interfaces are related to a concept of interaction that is not exploited by most current research technologies of human-computer interface. As stated by Bruner [13], the traditional interaction with computer-mediated information is based primarily on iconic or symbolic knowledge, and not on enactive knowledge. While in the symbolic form of learning knowledge is stored as words, mathematical symbols or other symbol systems, with the iconic stage knowledge is stored in the form of visual images, such as diagrams and illustrations that can accompany verbal information. On the other hand, enactive knowledge is a form of knowledge based on the active use of the hand and body for tasks of apprehension.

It is possible to say, then, from the discussions made in the present work, that enactive knowledge is not simply multisensory mediated, but that is somehow stored in the form of motor feedback and acquired by the act of “doing”. A typical example of enactive knowledge is established by the competence required by tasks such as typing, driving a car, dancing, playing a musical instrument, modeling clay objects, which would be difficult to describe in an iconic or symbolic way. This type of knowledge transmission can be considered the most direct, in the sense that it is natural and intuitive, since it is based on experience and perceptual responses to the motor acts [12].

Although there is a vast literature willing to explore the interaction between users and computers, enactive interaction is still an under-researched topic. In fact, research in Computer Science area are exploring some principles of this theory, however, often without realizing the implications they bring to the knowledge construction and possible paradigm shifts in the way people learn.

Thus, the control based processes on computers that are related to this type of interaction now require faster computers and systems able to handle more complex information (area in which researches are evolving rapidly), and new types of interfaces, computing architectures and software modules able to work with users in a complex way in terms of representations of information.

III. COGNITION IN MULTIMODALITY CONTEXT

Multimodal interfaces are known as inherently flexible, and to provide an especially ideal interface for accommodating both the changing demands on technology and the large individual differences present in the population [11]. These interfaces can be designed to support simultaneous use of input modes, to permit switching among them to take advantage of the modality best suited for a task, environment, or user capabilities, or to “translate” information from one mode to another in order to expand accessibility for users with selective limitations.

Dumas [7] says that the advantages of multimodal interface design are elucidated in the theory of cognitive psychology, as well as human-computer interaction studies. Findings in cognitive psychology reveal:

- humans are able to process modalities partially independently and, thus, presenting information with multiple modalities increases human working memory;
- humans tend to reproduce interpersonal interaction patterns during multimodal interaction with a system;
- human performance is improved when interacting multimodally due to the way human perception, communication, and memory function.

One area of cognitive research lies on how different modes of representation or action might influence, or shape, human interaction, attention, perception, interpretation, and meaning making. Researches suggest that different modes of representation (visual, audio or tactile), interaction (gestural, physical, visual) and communication change the way that information is perceived, and understood [14].

Another line of cognitive research focus on users’ active role in the interaction with multimodal messages [13]. It investigates general and individual patterns in reception of multimodality, particularly factors influencing the perception and interpretation process, such as users’
interests, attitudes, goals and motives, prior knowledge, experiences, individual preferences, emotions and expertise.

IV. MULTIMODAL INTERFACES UNDER THE PERSPECTIVE OF ENACTION

The human beings interact, in their daily life, with a wide range of digital situations that enhance their habits. Devices and artificial mechanisms are part of their daily lives, increasingly and in the most basic activities. That seems to be the result of a post-modern ideology, in which the demand for the new and original resources (both physical and signical), open precedents for the great use of media resources in the figure of several appliances.

Nevertheless, it is possible to see in our daily life, at work, at home and at leisure time, a narrowing of the relationship between individuals and the glimpse of McLuhan [15] about technology: it is the extension of man. This theory, already in 1964, said that with the arrival of electronic technology, men had elongated, put out of themselves a living model of your central nervous system. He pointed to the fact that all means are extensions of some human faculty, mental or physical (the wheel an extension of the foot; the book, of the eye; clothing, of the skin). Throughout the facility of the extended daily routine raised a new way of creating and thinking the communication as a function of the cognitive aspects that modern man has.

The man now works with perception in a sensory and synaesthetic way that goes beyond the information received, but considers the environment and the information contained therein. This man, contemporary user, enlarges himself with this technological mediation natural from a dynamic medium that adapts to each new experience, creating new patterns, new insights and generating new features. Thus, the mutant energy of experiments approaches increasingly the speed of thought, effectively turning the information access devices and interaction with the environment and the people who inhabit it, an extension of the human body.

It is then possible to explain the fact that the popularization of portable digital devices and the trend that they come to be increasingly coupled to man, namely, the information about the environment becomes part of the user (not more of the device), becoming attached to the means. To Kerckhove [16], the digital mean (electronic) is understood as an organ of control and analysis of the social body, since it is subject to social changes once the computer networks spread information and habits not only seen tracional means but also on several objects-prosthesis networked.

The contemporary digitalism, the intensive use and, at some point, the dependence on computers in their various forms, broke and at the same time created new paradigms. The idea of the global network has promoted great possibilities, including personal and private ones. The knowledge and experience in this new environment reveal mimicry of man to his creations. The expansion of the potential of human activity, mainly through new media devices, has resulted in new ways of conceiving the thought.

In an essentially technological contemporary scene, the trend of readings through technological artifacts allows a constant migration from real to virtual and vice versa, leading to a new situation on daily life, routines, actions and tasks, and especially on perception. In this sense, it is possible to say that from the popularization of technologies and technological resources, these have become more than an extension of man's body but also your mind.

Habits acquired from digital daily routine arise in simpler ways: in phones and smartphones, in MP3 players, tablets in the hands of the waiters, in the children's play. As in a vicious circle, every day, digital resources are produced, supplying desires or even creating new needs, presenting the user - of all ages and social classes - a new dialogue that develops new forms of communication, sometimes even new languages.

Thus, it is possible to see some movement in the industry, and especially in academia, towards the development of research and products that consider new interfaces and interaction paradigms, in which often happens to be considered the possibility to work in an intangible way, with interfaces merged to the organism which it interacts.

The aim of this article is to consider Varela’s theory [9] in favor of an approach that moves from fact to consider the body and its capabilities as something closely connected to human cognition. Maturana and Varela [2] describe that the term enactivism suggests that cognition depends on a dynamic set of relations and associations dependent of the context.

For Maturana and Varela [2], to face the problem of understanding how the experience, that they call praxis of living, is coupled to the surrounding world, full of regularities, passes by the perception that these are the result of biological and social stories of people.

Victor [17] uses the concept of amplifying human capabilities provided by the tool, for example a tool that increases the man can do, providing the ability to do what you want. The author does not speak specifically about human needs, nor about technology, he focus his speech on human nature. According to him, human capabilities are neglected throughout the process of designing new products and interactions: do not think about what people can do.

Researches carried out by several authors, notably the pioneering Leroi-Gourhan [18] and Napier [19], demonstrated how, in the evolution of Homo sapiens, the physical opposition between the thumb and fingers was becoming increasingly more articulated, associating with
subtle changes in the bones that support and strengthen the finger.

Leroi-Gourhan [18] studied the Mediterranean societies’ technique using the concept of basic means of action on the subject that is based on hands. In his classification, gestures may be those of: suspending with the fingers, pinching between the fingers, grasping with the whole hand and finally contain cupped hands.

Victor [17] highlights that man lives in a three-dimensional world. Human hands are designed to move and rotate objects in three dimensions, to pick up objects and put them over, under, beside, and within each other. No creature on earth has dexterity comparable to this.

In a direction often different from that presented in the critique of Victor [17], however, is the Academy. As an example of such divergence is the work of Dr. Todd Coleman [20], associate professor of bioengineering and director of the Laboratory of Neural Interaction. He was recently been recognized for its innovations in the development of flexible electronic sensors that adhere to the skin and transmit wireless signals to applications in brain-machine interfaces, monitoring of pregnancy, and monitoring of the newborn brain.

A brain-computer interface is a system comprising a means of direct communication between the brain and an external device. Coleman’s research group has developed an interpretation of BCI (Brain-Computer Interfaces and Brain-Machine Interfaces - BMI) as a system composed of multiple agents cooperating to achieve a common goal. This “theory decided in teams” allowed them to leverage ideas on control and information feedback to develop brain systems of direct control, which are easy to use, are theoretically conceivable, and reach a performance not previously attainable. This type of interface represents great improvement in the area.

Such interaction paradigm has three main components: a sensor that measures the neural activity of the user, an algorithm that maps these measures to control signals and send them to the external device, and a mechanism that provides feedback to the user about the resulting state of device. Measurements can be from sensors as a non-invasive electroencephalogram (EEG), which observes the gross electrical activity of neurons, or they may be from more invasive sensors as an electrocorticography or intracortical electrodes, often placed in the primary cortex, which may observe sets of neurons or their individual activity [21].

The feedback can be provided by a graphical representation, by vibrating matrices, by physical coupling with the external device, or by direct cortical stimulation. BCI has been used to control a growing variety of external devices, including computer cursors, text reading devices (spellers), artificial limbs, humanoid robots and wheelchairs [21].

Depending on the breadth of the definition of wearable computing, the first wearable computer may have been an abacus hanging around the neck on a string, or worn on the finger [22]. More recently came the electronic calculators (which could be carried in a pocket or wrist), and electronic watches.

A common understanding of the term "computer" is that it is something programmable by the user while being used, or it is relatively general in nature (eg, the user may change the programs and run many applications).

In the 1970s and early 1980s, Steve Mann [22] designed and built a series of laptops, equipped with systems for several purposes, which include several kinds of sensing, biofeedback, and multimedia features, wearable musical instruments, audio-based computers and assistive devices for the blind.

Humans evolved and acquired, for example, the great thing of visual (rapid eye movements that aim to put in the retinal foveal the image of a visual target appeared in the periphery of the retina), and without walking or moving his head, people receive in their system of visual perception, the movement data. Thus, the enactive approach draws attention to the fact that hardly perception (and conscience) can be isolated in only one side of the relationship established via interface. It may be important to include more of the system based only on the relationship between mind and environment.

V. DISCUSSION

The interdisciplinary field of cognitive sciences has traditionally been addressing issues to explain how our cognition is modeled and how our understanding of the world is built. There are, however, other notions of computing that figured in the histories of both the computer and the cognitive sciences. The externalist view currently explored among scientists in cognitive science refers to the theory of Embodiment and Eaction and they were strongly developed and researched by Varela et al [3].

For over half a century in philosophy, and about 20 years in Artificial Intelligence and related disciplines, there has been a reassessment of the nature of cognition. Instead of emphasizing formal operations on abstract symbols, the new approach focuses attention on the fact that thinking occurs in very specific (and often very complex) environments, is used for very practical purposes and explores the possibility of interaction with manipulation of external media. Thus, foregrounds the fact that cognition is and embodied or situated activity, and suggests that thinking beings ought therefore to be considered first and foremost as beings that act (do).
An enactive approach for interaction design, searches interfaces that emphasize the use of multiple senses and must then provide an appropriate framework for addressing possibilities of technological interfaces mediate the relationship between users and the world, so that they can improve perceptual interactions. An enactive approach should consider the sensory "input" and the motor "output" as two facets of the same process of generating meaning.

One can also say that the design of interfaces should be targeted to expand the range of possible interactions between people and the world through technological devices, and can therefore be referred to as a magnification factor of sense making. Such expansion is to provide opportunities for people to create new modes or forms of perceptual interactions. Thus, a mode of perception is perceived as skillful negotiation of a specific relationship with the environment that allows significant access to a certain class of aspects of the world, directly affecting the experience of the subject [11].

Enactive interfaces are driven by the interaction, while the users actions become the means to the system to provide an appropriate response through the immediate sensorial experience. These systems work through the conception of human being’s natural cognitive process where the user can learn while acting without the need to read a manual. Some examples are grasping an object, ridding a bike and kicking a ball.

The role of the interface, in this case, is to promote complex relationships (experiences) and therefore, being a film that, unlike traditional approaches, is not between the person and the related object, but is impregnated of such experiential possibilities. Thus, there is a kind of skin that permeates the whole process of interaction that, because its fractal structure, enables a very large range of possibilities and extensions of that contact.

Non-enactive interfaces are designed only to provide information exploring the visual sense, presenting the information through illustrations, icons and others. On the other hand, innovative devices as wii and kinect provide user-system interactions acquiring, storing and interpreting their gestures as input information, that is the key of Enactive approach.

What you see, underlying representations of the world created by beings acting on it, their formulations in declarative sentences, for example, are no longer representations, but the right idea in the world that has as an agent that acts on it. In other words, using the term enaction aims to highlight the sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition.

Varela et al [3] argue that in-between amalgamating the structure of the world (while perceived as a way to life and created by means of certain organs of the body - enacting) and contextualized meaning given to experience (emergency concept), you can practice the attentive presence to the world.

Daily new products, technological or otherwise, are developed and commercialized, making it necessary to acquire new skills so you can deal with digital technology and its ways of interacting.

When proposing multimodal interfaces under the enactive perspective what happens is the opposite of User Centered Design that promotes sedimentation of constraints common to some groups of users without a real concern to understand the interactive possibilities that their actions in the world and their technological devices can provide. In the Multimodal Enactive approach, it is possible to perceive the real value of individualities: it must activate the objects that belong to the universe from every direction, producing multiple sensory environments and rich in sensations that arise regardless of the fullness of the senses and enable a world created from the imagery of the individuals working in it.

VI. FINAL CONSIDERATIONS

Cognitive infocommunications (CogInfoCom) study the connection between infocommunications and cognitive experiences. This paper explores Enactive knowledge and its relations with Interaction Design of Multimodal Interfaces. Enactive knowledge is what is built through the action and is built on motor skills as manipulating an object, riding a bike or practicing sports. It is acquired by doing.

It is clear that, in fact, there is little explicit discussion about the meaning of enaction. Thus, when considered in the abstract, particularly as an indication of the focusing area that contrasts with the interest in studying the symbols, this already sets interesting field of research. But when you have the attention focused on a concern about the possibility of realization, the research becomes even more relevant with claim to significantly contribute to the development of the areas involved, especially the interaction design.

In the enactive approach, the environment is considered as a determining factor and added to the body. Thus, consider other perceptual channels in the construction of interfaces providing differentiated experiences and multisensory, is a trend of research that considers externalist theories of philosophy, and specifically enaction as a paradigm. Thus, a line of research and development has been gaining strength over the last decade: the enactive interfaces.

Enactive interfaces works with the users natural process to interact with systems, in a sense that the systems may interpret users’ gestures providing responses in terms of cognitive aspect. Users learn with experience while systems become more intuitive and perceptual.
Multimodal interfaces are candidates for the creation of enactive interfaces due to the coordinated use of sound, touch, haptic and vision. In fact, there is no simple answer to a challenge as complex. This research aims to arouse the interest of researchers in Interaction Design in researching these movements relatively unexplored Cognitive Sciences.

Multimodal interfaces are known as inherently flexible, and to provide an especially ideal interface for accommodating both the changing demands on technology and the large individual differences present in the population.

This research contributes to the main goal of the Enactive Network of Excellence, a European consortium with more than twenty research laboratories that are gathering resource efforts to define, develop and explore interfaces of this nature.

REFERENCES