Perception of Virtual Multi-sensory Objects : Some Musings on the Enactive Approach

Annie Luciani[1], Sile O'Modhrain[2], Charlotte Magnusson[3], Jean-Loup Florens[1], Damien Couroussé[1]
[1] ICA laboratory and ACROE, Grenoble Institute of Technology, France
[2] Sonic Arts Research Centre, Queens University of Belfast, Ireland
[3] CERTEC, University of LUND, Sweden

Abstract

In this paper we explore, by means of three pilot observational studies using virtual objects, how direct perception through action of multi-sensory audio-visual and haptic object properties support the creation of new categories of believable and plausible objects than can be perceived as being different from those that were presented The three experiments are based on variations of "Pebble boxes" and consist in the exploration and the manipulation of multiple moving multi-sensory objects (the Pebbles). Results from observations and informal interviews with participants illustrate how an inferred scene is apparently constructed from experience, as assumed in the cognitive Enactive concept, by means of three complementary strategies: the Emergent Exploratory Procedures, Dynamic Manipulation Adaptation, and Adaptive Experimental Learning. Findings also illustrate the complementarities between the so-called ergotic and semiotic situations with respect to the strategies that were apparently successful in inferring a believable and plausible scene.

Several fundamental questions arise which are relevant to the Enactive assumption with respect to the coupling between perceiving and acting, some of which are discussed here.

1. Introduction

The work presented here lies in the intersection between two open questions:

• What variables are tracked by the haptic and auditory senses in the perception of multiple mobile objects?

• How do these variables function in differentiating between the perception of virtual objects and that of real objects altered by digital processing?

In the field of perception of numerosities, a wide quantity of work exists concerning the numbering of perceptual stimuli: judgments on numerosities in visual stimuli, counting of acoustical stimuli and more recently, of unimodal tactile and bimodal tactile/visual stimuli [1] [2].

In the field of virtual reality, most research is centered around performing comparative experiments in order to implement virtual reality platforms that are as similar as possible to a real referent situation such as when using VR for the training of manual tasks or when evaluating perceptual paradoxes. Several others aim at investigating perceptual multi-sensory cues in an environment that is easier to construct than a physical environment. This very new field focuses on such issues as the study of cross-modal discrepancies in object perception [3], [4].

However, new questions are raised by the common use of virtual objects and their multi-sensory manipulation. Virtual objects – i.e. things for which the behaviors are computed by digital machines, sensorially transduced by Digital/Analog transducers (output devices) and acted upon by users through Analog/Digital transducers (i.e. input devices) must still be considered as "real objects" in the sense that they are really present for our perceptual systems and really affected by our actions.

The question of the perception – and more generally the identification and the appraisal – of such "strange" real – albeit virtual - objects is a topic that is now attracting more interest, particularly in the realm of research on the believability of virtual constructs such as virtual objects.

The work presented here addresses this question by asking how we can identify the existence of numerous moving objects from the perception of numerous stimuli? There is no predetermined relation between the fact that there are numerous objects and that we

978-0-7695-3381-0/08 $25.00 © 2008 IEEE
DOI 10.1109/CW.2008.107

perceive numerous stimuli. For example, numerous auditory or haptic events (such as impacts) can be produced by only one moving object (clapping your hands or hitting a surface producing several auditory events). In the absence of vision, it is not easy to determine if a series of perceptual events (auditory or haptic) are produced by one or several moving objects. Perceptual numerosity judgments are not necessarily equivalent to the actual number of objects present. The question is then: from what types of perceived features do we distinguish a single source of multiple stimuli from multiple sources of single object-related events.

This question is of crucial importance in the cognitive reconstruction of virtual objects as it addresses a basic question of cognition: What proportion of a constructed scene is related to the direct perception of object properties and what proportion is due to more abstract object-related percepts? Can a given set of percepts, under different perceiving conditions, infer the presence of at one time a collection of simple objects and at another the presence of a single complex object? Moreover, can a set of conditions be created which will invoke one situation rather than the other? Finally, in what conditions for rendering virtual objects are the modalities – vision, audition and haptics – coherent and when are they discrepant?

Our approach here has been to begin this process of exploration by conducting a series of qualitative informal studies that can then guide the design of more conventional quantitative experiments.

Three important aspects of the current study have been previously explored in [5] using a) a physical, tangible pebble box controlling virtual sounds and b) an entirely virtual pebble box:

(1) Estimation of the number of objects: participants were asked to assess the number of objects inside a box.

(2) Sensorial preferences: Participants were presented with different auditory feedback and asked about believability and likeability of such virtual scenes.

(3) Sense of control: Participants were asked to assess how much in control of the sound they considered themselves to be.

Other aspects, mainly focused on the perception of audio-haptic complex virtual environment have been also previously explored in [6].

The work presented here extends this research in the field of perception and identification of multiple moving objects, by exploring complementary questions directly tied to the concept of enaction:

1: What is the scene (or object) inferred from the sensory experience, i.e. from the performed action and sensory feedback (auditory, visual and haptic)?

2: Are all three sensory modalities participating in the creation of the evoked scene, and, if not, which modalities are not and why?

The current paper presents three successive tests using three different realizations of pebble boxes, and presents the main results of our observations. It concludes by presenting a comparison between the results obtained using each platform with respect to the enactive paradigm and the ergotic / non-ergotic properties of the situation [7].

The experiments on all three platforms were performed by 10 participants ranging in age from 25 to 55, 4 of whom are female. 6 of the participants are professional musicians and one is congenitally blind. All the observation sessions were recorded on video with audio capture of comments.

2. Platform 1, the physical "Pebble Box"

2.1. Description of the platform and of the experiments

This platform, designed by [8] and used in [5], consists of a real pebble box used as a tangible interface to control synthetic sounds. A number of pebbles (approx 400) are placed in a box (Figure 1). The sounds produced by the real pebble collisions are picked up by a microphone and analyzed in software to extract parameters such as event dimming, amplitude etc. These parameters are passed to a granular synthesis algorithm which applies the values as parametric controls of a stored sound sample or 'grain such as a water drop, a bird call, an apple being crunched, and so on. During the test the participant is invited to freely manipulate the pebbles.

For the current investigation, three configurations of the Pebble Box were used. In each case the haptic environment consisted of some number of pebbles while the auditory environment produced the sound of water, birds or eating apples respectively.

The questions the participants were asked to answer were:

Q1.1. What are the scenes this experiment suggests?

Q1.2. Does the association of the sound and the virtual haptic objects you are manipulating seem, with (A) believable, (B) artificial, (C) not sure.

Figure 1. Platform 1: "The tangible Pebble box"

As in [5] the believability ratings were done totally subjectively, and with each person's own notion of believability. It is an interesting task to explore more in the future how the believability notion for this type of environments could be developed.

2.2. Results

Objectively, the environments being explored are paradoxical in the sense that they don't correspond to previously known scenes: the pebbles you can feel produce sounds that are not pebble sounds.

2.2.1. The Inferred Scene

In all three cases, i.e. water, birds and apples, all participants appear to have constructed some representation of the environment they had explored that did not entirely correspond to the presented environment but always reflected some aspects of this environment. The following comments are extracted from video recordings of participants as they explored the three environments.

Bird sounds

People imagine "walking on a gravel path or throwing a stone, triggering panic among birds nesting in nearby bushes.

Water sounds

People imagine "handling stone(s) in water or disturbing animals (fishes) which escape".

Apple eating sounds

People imagine "an animal within the box" and become anxious.

2.2.2. Emergent Exploration Procedures:

The inference of a possible scene is a dynamically evolving process, alternating between hypothetical scene imagination and changes in strategies for exploring the environment. From our observations, these two 'threads' appear to converge as a plausible explanation is arrived at. Inspired by the work by [9] we refer to the hand movement strategies that result as "Emergent Exploratory Procedures" (EEPs). Some of these EEPs are mimetic in nature (fingers running amongst the pebbles) while others are more metaphoric (cautious tapping on the outside of the box to attract the attention of the animal eating the apple in there).

2.2.3. Believability:

First, in line with the heuristic observation reported in [5], participants consider that the situation when they are manipulating pebbles and hearing the sound of water to be the most likely or plausible. Perhaps the most surprising finding is that none of the participants considered any of the inferred scenes they created as they explored to be implausible. Even when the correlation between tangible and audible feedback was highly unlikely to arise in a real-world situation such as the bird call or apple eating, participants are surprised, but pleasantly so (see photographs below).

Finally, it is not surprising that they modify their manipulation strategies, as it seems they do so to reinforce their interpretation of the situation.

Figure 2. Photographs of the experiments on Pebble box 1 – Left : "surprise but finally believable"; right : "the participant mimics walking on a gravel path to disturb ("synthetic") birds".

3. Platform "Pebble Box 2"

3.1. Description of the platform and of the experiments

The second platform is comprised of a PHANToM haptic device connected to a multimodal simulation of virtual pebbles (see Essl et al,2005, for a full description). Using the PHANToM's pen-shaped manipulandum, participants move in an environment containing 10 physically-based 3D haptic models of either cubes or spheres moving in a squared box, with the dynamics simulated by the Open Dynamics Engine software (Figure 3). Sounds of collisions are triggered by means of a collision detection algorithm. There are also a number of options for presenting visual information about the environment:

- ⇒ .The 10 objects can be made visible or invisible.
- ⇒ The position of the PhanToM cursor is visible or invisible.
- ⇒ The simulated pebbles can be virtual cubes or spheres.
- ⇒ Haptic cubes can be represented as visual spheres and vice versa.
- ⇒ Visual size can be greater, smaller or equal to haptic size.

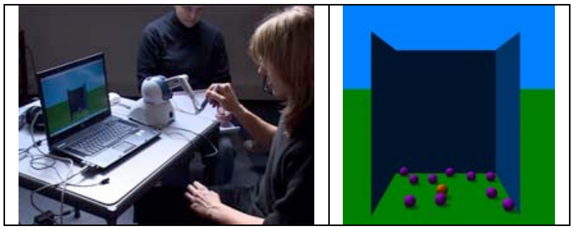

Figure 3. Pebble Box platform 2: VR Haptic manipulation & “Changing the visualization”

3.2. Results

In a similar way to the Pebble Box Platform 1, people try to infer a believable scene, if possible from all the multi-sensory feedback they are receiving in order to create for themselves a plausible scene. Again this inferred scene may be quite different from that constructed by the VR environment designer. Very often, the inferred scene is far from the implemented scene (in this case ten rigid pebbles moving in a box):

(1) When no visualization is present, people tend to infer a phenomenon such as a force field or magnetic field rather than « perceiving clearly defined objects such as pebbles ». This also extends to the cases involving visual feedback. For example:

When visual objects are smaller than their haptic objects (as defined by the physical model from which they are both derived)), the inferred objects are believed to be surrounded by a transparent shell or extended by a force field. Larger visual objects result in an illusion of a soft surface.

(2) Subjects change the way they manipulate objects in the VE when they can see the position of their notional fingertip by means of the virtual cursor. Again we observed here the emergence of certain exploration strategies, i.e. EEPs. For example:

- *When they can see the 3D cursor, participants tend to explore the whole space, whereas without this feedback they do not.*

- *They attempt to create the conditions allowing them to explore the shape of the supposed objects.*

Unlike the case of the physical Pebble Box Platform 1, when participants cannot easily infer a plausible scene involving all the perceptual modalities, they are lead to suspect that there is an inconsistency in the rendering of the VE, sometimes distrusting the haptic, sometimes the auditory and sometimes the visual feedback.

4. Platform “Pebble Box 3”

4.1. Description of the platform and of the experiments

The third Pebble box platform is a 2D virtual Pebble Box composed of a circular box containing 8 mobile masses whose apparent rigidity can be controlled and hence which exhibit more or less elastic behavior when they collide. In addition the tip of the manipulandum is virtually coupled to a rigid mass whose rigidity can also be controlled.

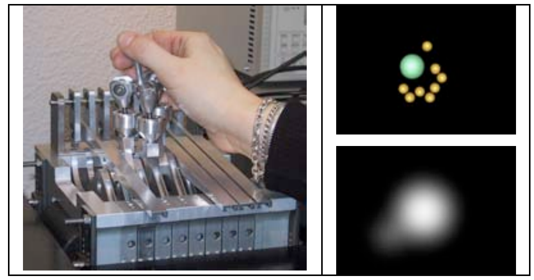

Figure 5. Pebble Box platform 2: VR Haptic manipulation & “Changing object shape & visualization”

By changing the physical parameters of the interaction between each pair of virtual pebble-masses and between each pebble-mass and the haptic stick-mass, we modeled the changes in the VE as follows:

⇒ Elasticity / rigidity : high, medium, low

⇒ Viscosity: high, medium, low

⇒ Interaction radius between the haptic stick and the pebble-masses: big, small.

Two visualizations have been tested:

⇒ Visualization 1: a ball like visualization (see figure 5, right, upper)

⇒ Visualization 2 : a blurred medium-like visualization (Figure 5, right, lower)

The sounds are the sounds produced by the simulation of the pebbles at acoustical frame rate (44KHz). When colliding, the two pebbles are vibrating in a physically coherent way producing one sound for each collision, as would be the case in the real world. The sound signals generated depend on the physics of the collision (material properties of the colliding objects, strength of the impact, object velocities).Thus, this simulation is objectively a very realistic physical simulation of identical pebbles colliding and being pushed by another object.

A first series of observations was carried out [10] in which 6 participants manipulated 54 different models of this virtual Pebble Box, that were: 9 physical parameter sets for the Pebble Box always containing 8 masses * 6 different sensory feedback conditions. From these preliminary 54 experiments, in order to reduce the number of experimental cases to be performed by our current 10 participants we extracted 4 well-categorized cases, unanimously selected by the earlier 6 participants. These four cases are (Table 1):

Table 1. Experimental cases

Parameter	1	2	3	4
Pebble rigidity	high	low	high	high
Stick-pebble rigidity	high	low	high	high
Viscosity	medium	medium	Low	Low
Pebble size	Big	medium	medium	small

4.2. Results

As with Pebble Box Platforms 1 and 2, all participants tried to infer a plausible explanation for what they were perceiving. If they could not always find a plausible explanation, as with Pebble Box 2 they began to distrust the consistency of the multimodal rendering of the VE. As before, the inferred scene is not necessarily coherent with the objective simulated one, so that again it appears participants construct for themselves a plausible explanation that can account for the relationship between the movements they make and the feedback they receive.

(1) The value of physical material interactions leads participants to infer two types of very different categories of scene not necessarily similar to the objective one. For example:

- *People feel a kind of « medium », « paste, « force field », « cotton », etc. when the virtual masses (or virtual pebbles) are in soft colliding interactions.*
- *People feel clearly defined objects but not necessarily all of them.*
- *People feel clearly defined objects but not of the same size.*
- *When the sound or the visual feedback are not consistent, people tend to infer a more 'correct' form of the feedback, some 'correcting' the sound and others the visual cues.*

(2) The dynamic of the coupling of the manipulation, for example the type of grasp used to hold the ERGOS manipulandum [11] by the participant appears to be adjusted depending on the character of the VE being rendered. People appear to adapt the dynamics of their moving arm and the dynamic of their grasp to account for the dynamic properties of the simulation. This can be interpreted as a kind of Dynamic Manipulation Adaptation and suggests an evolving strategy for acting in the VE as the properties of the environment are perceived, through action, and 'learned'. For example:

- *When objects are in a strong rigid interaction, people grasp the device firmly and act (press, move) with high energy.*
- *When objects are in soft elastic interaction, or when they are very small, people manipulate more delicately for example by grasping the stick with finger tips only.*

Figure 6: Dynamic Manipulation Adaptation. Left: Manipulation of strong rigid interaction. Right: Manipulation of soft elastic interaction

(3) The exploration of the environment becomes more refined. Throughout our observations we noticed that the scene inferred was progressively changing in correlation with the refinement of the reported sensations. This can be described as an adaptive learning of the experimental environments. For example:

- *When the simulated matter is very soft, people start by reporting that they feel nothing and progressively tend to conclude that they feel a type of « resistant or viscous » field, or field + of lumps.*
- *When the simulated matter is very rigid and the objects very big, without visualization, people start out by feeling « one big object », explore its shape and progressively discover other objects. They then imagine these to be smaller than they really are.*

5. Comparison between the three platforms and their observations

5.1. Enactive experience

In all the observations, with all the participants, a common factor that appears to persist is that an inferred scene which is constructed from the sensory-motor experience. This inferred (constructed) scene can be different from the objective scene as conceived by the designer. Thus we can talk about the act of creating a scene for our participants to explore, but what we have actually done is more to assist the cognitive creation of objects than to support the identification of objectively rendered entities.

Using virtual objects to explore how real objects are emerging as cognitive categories, we can say that these observations, because they take place in such a free experimental contexts, provide a means of probing the enactive paradigm: *"...cognition is not the representation of a pre-given world by a pre-given mind but is rather the enactment of a world and a mind* *on the basis of a history of the variety of actions that a being in the world performs." [12].*

During this process of inference of a plausible scene, participants appear to have developed three complementary strategies:

• Emergent Exploratory Procedures (EEP), strategies which they are seeking out, on-line, and which are emergent as assumptions about the coupling between actions performed and feedback received are being tested. These exploratory procedures may use different modalities: mimetic, metaphoric, dynamic adaptation, sensory reinforcement, etc.

• Dynamic Manipulation Adaptation (DMA) in which participants adapt very quickly and "on the fly" their manual dynamic performance to account for changes in the sensory feedback and the imagined felt objects. In this way exploratory procedures are used as a means of converging efficiently upon the most plausible scene.

• Adaptive Experimental Learning (AEL): is more an associated effect than a structural means of creating objective categories. However, we observe that people learn very quickly dynamically and "on the fly" what are the best manipulations to allow them converge quickly to a plausible and stable scene. This is a derivative property of intimacy and embodiment, supported by these types of experiences. Subjects are "with" the object, that leads to one sense of presence: besides the sense of "being there", they develop the sense of "being with" the object, of sharing its environment.

Consequently, and in contrast to our thinking at the outset of these observations, participants never believed the objects that we presented were totally implausible. Moreover, when the sensory feedback received as a consequence of an action cannot be integrated easily in the expected sensory experience, participants do not question their own perception, but reject any such inconsistencies as artifacts: *a sign of the virtual nature of the environment, a discrepancy with respect to reality*. This observation could be compared with the phenomenon of the uncanny valley in the perception of virtual human-like faces, in which a discontinuity appears when the resemblance to real faces increases.

5.2. Ergotic / non-ergotic interactions

One of the reasons for performing these investigations on the three platforms is their complementarity regarding the physical consistency between gestures and sensory feedback. An efficient typology for such an investigation is the typology introduced by Claude Cadoz [7], who distinguished ergotic interactive situations, and non – ergotic, purely epistemic and/or semiotic situations. According to this typology, an ergotic interactive situation is characterized by the fact that there is a physical energetic coherence along the whole interaction chain, from the originating action to the sound and the image that are produced by the action, as is the case in musical instrumental performance. As claimed by these authors, this property of "ergoticity" is a fundamental property to support intimacy and embodiment [13]. In non-ergotic interactive situations, such as the mouse actions for a WIMP interface, sign language, musical conducting, etc., the information exchanged by the two interacting bodies is purely semiotic (provided by the 'actor' to the environment) or epistemic (received by the actor from the environment) or a combination of both, without physical energy being exchanged between them. We may thus compare the three observation situations with respect to Cadoz's typology of interaction functions :

• In Pebble Box platform 1, the relation between gestures (applied to the real Pebble Box) and the auditory sensory feedback is purely semiotic / epistemic. Due to the computer process that maps the gestures onto synthetic sounds, there is no possibility to imagine a physical direct relation between the manipulation of the pebbles and the sounds (the birds sounds for example). The relation takes place only at high symbolic cognitive level.

• In Pebble Box platform 2, the relation between action and visual feedback is totally managed by physical modeling and so it is an instrumental ergotic relation in which there is a physical consistency between the actions performed on the force feedback device and the movements of the visual objects. Conversely the sounds are prerecorded sounds triggered from a signal event and consequently the audio feedback corresponds to a non-ergotic situation.

• In Pebble Box platform 3, the relation between action and all the feedback conditions is totally managed consistently by physical modeling and so it is an instrumental multi-sensory ergotic relation in which auditory, visual and haptic evolutions are physically linked and coherent.

Two remarkable observations can be extracted from the observation sessions:

1. This situation is more ergotic, and the effects of dynamic adaptation and dynamic on-line learning are more pronounced. Less strong are percepts such as believability of the situation, in the sense that some modalities are sometimes distrusted. This is very noticeable in Pebble Box 3. Despite the fact that the sounds are "objectively" very realistic (two pebbles produce sounds in a more physically realistic manner, they cannot be easily associated with the idea of pebble

collisions. This is probably due to the very high number of impact sounds, that is equal to 2 times the number of collisions. Sounds tend to be considered as noisy and non easily interpretable.

2. The less the situation is ergotic, the more important are the emergent exploratory procedures and the greater the reliance on metaphor. In addition, most of the scenes are believable in accordance with the level of coherence of the sensory feedback. We observed that this believability level is less modified when people perceive the two auditory feedback sources simultaneously: the real sounds of the real pebbles and the synthetic sounds. As noticed, they imagine moving pebbles in water, hearing the sounds of the pebbles and their effects in the water even though their hands remained dry.

6. Conclusions

We have conducted qualitative observations in order to initiate new research on how virtual objects might be perceived, assuming that virtual objects are also in and of themselves new real objects that can be acted upon and explored in a stable context. Such research differs from other work using virtual objects to increase our understanding of perception or research that uses VR platforms for training because we place participants in a situation there the world they explore has no basis in reality – it is a novel multisensory environment that they must learn about by acting upon it and thereby discovering how it operates. In this way is becomes a “laboratory to capture how humans construct the notion of what an object is”. Thus, the method is to analyze “natural human behaviors and spontaneous and free comments”, when they are discovering through their action and perception, the properties of a virtually constructed world.

We have observed new features such as: the emergence of exploratory procedures, dynamic adaptation on the fly, iterative process of dynamic learning, all of which can then be explored in quantitative follow-up studies. But above all, we can conclude that such experiments lead us to a greater understanding of different assumptions about the coupling between perception and action and could thereby help us to realize practically the theoretical implications of the Enactive approach – to build virtual manipulable objects that can be used as tools in virtual environments or as virtual tools to act upon the real environment. From our observations we might expect the real properties of such tools to emerge from the cognitive experience of their operators and not only by extrapolating them from the properties of the real objects that served as a first reference in the design of the virtual objects.

7. References

[1] Gallace A, Tan H Z, Spence C, 2006, "Numerosity judgments for tactile stimuli distributed over the body surface" Perception 35(2) 247 – 266

[2] Gallace A, Tan H Z, Spence C,, "Multi-sensory numerosity judgments for visual and tactile stimuli," to appear in Perception & Psychophysics, Vol. 69, No. 4, 2007.

[3] Woods, A., O'Modhrain, S. & Newell, F. "The effect of temporal delay and spatial differences on cross-modal object recognition." Cognitive, Affective and Behavioral Neuroscience, Vol 4, No.1 PP260-269, Psychonomic Society Publications, June 2004.

[4] Ernst, M. and Bülthoff, H. Merging the Senses into a Robust Percept. TRENDS in Cognitive Sciences Vol.8 No.4 April 2004.

[5] Georg Essl, Charlotte Magnusson, Joakim Eriksson, Sile O'Modhrain. Towards evaluation of performance, control of preference in physical and virtual sensorimotor integration. Enactive 05. 2nd Conference on Enactive Interfaces. Genoa. November 2005

[6] Charlotte Magnusson, Annie Luciani, Damien Couroussé, Roy Davies, Jean-Loup Florens. Preliminary test in a complex virtual dynamic haptic audio environment. 2nd Enactive Workshop May 25-27 2006 - McGill University, Canada

[7] Cadoz C., Wanderley M. (2000). “Gesture-Music” – in “Trends in Gestural Control of Music”, M. Wanderley and M. Battier, eds, pp. 71-94 -2000

[8] O'Modhrain M. S. , Essl G. “PebbleBox and CrumbleBag: Tactile Interfaces for Granular Synthesis”. NIME 2004: 74-79.

[9] S.J. Lederman, R.L. Klatzky, “Hand movements: A window into haptic object recognition”, Cogn. Psy. 19 (1987) 342-368.

[10] Maxime Houot, Annie Luciani, Jean-Loup Florens. “Perception of multiple moving objects through multisensory-haptic interaction: Is haptic so evident for physical object perception?”. Proc. of Enactive06 Conference. Montpellier. Nov. 2006

[11] J.L. Florens , A. Luciani, C. Cadoz, N. Castagné. ERGOS : Multi-degrees of Freedom and Versatile Force-Feedback Panoply. Proc. of EuroHaptics 2004. Munchen, Germany. pp. 356-360. 2004/06/05-07.

[12] Varela, Thompson, Rosch. The Embodied Mind . 1991.

[13] Essl G. and O’Modhrain S.. « An enactive approach to the design of new tangible musical instruments ». Organised sounds 11 (3). 285-296. 2006. Cambridge University Press.

8. Acknowledgements

This research work has been performed with the support of the FP6 NoE Enactive Interfaces and French Ministry of Culture (Direction of Music and Dance), and the French Ministry of Research.