5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013

Cultural awareness and personal customization of gestural commands using a shamanic interface

Leonel Morgado[a, b]*

[a]INESC TEC (formerly INESC Porto), Rua Dr. Roberto Frias, 378 4200-465 Porto, Portugal
[b]UTAD – University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal

Abstract

Gesture-based interfaces to control systems using full body or upper torso motions are becoming more common, particularly in the entertainment industry but increasingly in other domains. These interfaces may represent significant hurdles for individuals with motion or balance impairments. Even for non-impaired users, commands which are not readily represented by body mimicry require users to learn new non-intuitive gestures for adding layers of complexity in commands. In this paper, a reflection and clarification of these challenges is presented, based on some sample usage cases. Then a possibly solution path is presented, resorting to an idea originating in a science fiction novel, the shamanic interface imagined by Daniel Suarez. A reflection on its feasibility and on possible paths towards its creation is presented, as well as the possible impacts to render gesture-based interfaces more accessible to all and potentially easier to learn due to enhanced cultural awareness. The basic idea presented in this reflection paper is that the creation of a concern-separation layer between the gestures being executed by the users and their interpretation by computing systems can contribute both to the access by users with special needs and to all users in general, by enabling customized approaches to gesture-based control.

* Corresponding author. Tel.: +351 259350369; fax: +351 2589350356.
 E-mail address: leonelm@utad.pt.

1. A brief panorama of the current scenario in low-cost body-based interfaces

On September 2010, I was madly jumping up and down – and side-stepping – to control a rubber boat over roaring rapids. It was the first public presentation in Portugal of Microsoft's new interface, Kinect [1], and the chosen game was Kinect Adventures! [2]. This is actually a collection of games, and the specific game I was experiencing was River Rush, where one or two people need to control a rubber boat by side-stepping in sync and jumping, with a significant amount of physical activity, as demonstrated in Fig. 1, a public photo recording that moment (the author is in mid-jump in the center, on the left side a screen projection of the game is visible).

Fig. 1. Kinect presentation session at Videojogos 2010 conference
(photo from http://gaips.inesc-id.pt/videojogos2010/contactos/fotos2/files/page19-1028-full.jpg)

Impressive as the Kinect was in the home videogame industry, for enabling low-cost full-body interaction without requiring players to carry specific gadgets, it was only an arguably major point in a trend towards making motion-based controllers available to the wide public, which had been initiated years earlier, albeit with little acceptance initially [3]. The market success of the Wii console launched in 2006 by Nintendo, with its motion-based command, the Wii Remote (also known as "Wiimote") [4], prompted responses from major players in home videogame console market, leading Sony to release its PlayStation® Move command in 2009 and Microsoft the Kinect in 2010.

This accelerated introduction of motion-based systems control has steadily been encompassing other uses, not just gaming. Ongoing developments on availability of low-cost sensors and computation has seen other approaches to motion detection become ever more present in our daily lives, not just with these game controllers but also with body-worn sensors such as those found on most smart phones and beyond them [6]. At the time of writing, at least two novel low-cost devices have been announced for gesture-based control of computers (and other systems), and their marketing materials all focus on non-gaming usage aspects: gaming is

presented as but a use among many: I am referring to the LeapMotion controller [7], which proposes to detect precise hand gestures above it, and the Myo armband [8], which proposes to detect the electric activity of arm muscles to identify gestures (both are represented in Fig 2). And researchers have presented the WiSee system, which reportedly identifies a small number of wide gestures using Doppler shift of wireless signals [9].

Typically, the use of such devices implies mimicking a gesture directly. This can place problems for motion-impaired users or, when the gesture takes place while standing, for balance-impaired users. For instance, older adults experience various limitations in movement control, stamina, and reaction time; balance and gait are diminished; and motor learning itself rendered more difficult – aspects compounded with pathologies such as arthritis and others [10]. However, the usage challenges are not simply for users with motor impairments. The mimicking approach itself is limited in terms of the range of commands necessary for gesture-based computer control, as will be clarified in the upcoming section.
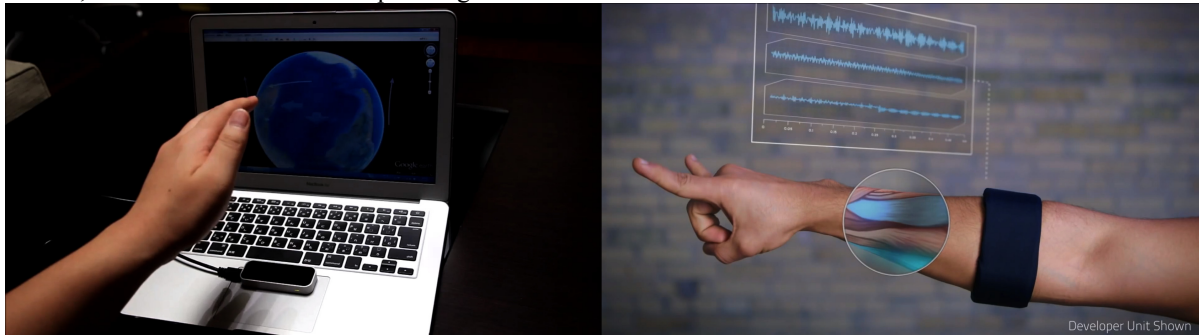

Fig. 2. (a) Leap Motion controller (on top of laptop trackpad); (b) Myo armband and superimposed schematics
Video stills from http://www.youtube.com/watch?v=cL3jCcQXmUc and https://www.youtube.com/watch?v=oWu9TFJjHaM#!

The basic idea presented in this reflection paper is that the creation of a concern-separation layer between the gestures being executed by the users and their interpretation by computing systems ("gestural commands") can contribute both to the access by users with special needs regarding motion and balance, and to all users in general, by enabling customized approaches to gesture-based control. This idea has some connections with that of "nomadic gestures" proposed by Vatavu [11], where each user would carry with him/her the gestures in a personal mobile device, which would be uploaded to the system for controlling it. However, the nomadic gestures approach would require standard gesture sets to be agreed upon by every software developer, which could then be replaced by the user-uploaded versions. Instead, I put forward the proposal that applications and systems should in their architecture separate gestures (motions in space and time) from gestural commands, just as keys on a keyboard are separated from the actual characters sent to applications on text input. As long as the identification of a gesture implies a conversion into an abstract gestural command, the opportunity lies for alternative conversions to be implemented, thus expanding the reach and access of gesture-based control. In short, a system's input settings should account not only for keyboard layouts and mouse laterality, but also for gesture converters.

2. Sample usage cases on the limitations of body mimicry

2.1. Difficulties due to precision and synchronicity

Execution of movements involves significant control over the precision of the motion of limbs and body. It is only natural that some of the earliest games for motion-based controllers were dance games, which provide an example of the limitations of mimicking. In these games, typically the player needs to imitate the

movements of a dancer monitor or leader, either for simple matching or for controlling another virtual dancer. Fig. 3 shows an example of a typical gameplay moment: the player does not need to provide the exact limb positions and timing in order to play: rather, the system tries to interpret the similarity between the player's motions and the intended ones, and the virtual dance team on the screen displays perfect synchronized moves.

It is common for the tolerance between intended and actual moves to be a difficulty setting. But this need for tolerance in the execution of movements is not a characteristic of dance games: all gesture interactions are by nature "analogue", not digital, and thus need to be converted to digital counterparts, a process which requires approximations, sampling, error tolerance, and other common concepts of analogue-to-digital conversions. This process may be further complicated by cultural variance in the execution of gestures, as the "Wave like an Egyptian" demonstrated by detecting cultural background from somatic behavior [21].

Fig. 3. Gameplay of Dance Central 3 using Kinect.
Video still from http://www.youtube.com/watch?v=vuiRUV5LWuk

Another example of the need to compensate for lack of precision in human motion is even more commonplace: pointing (as witnessed by anyone subjected to slide presenters dangling aimlessly a laser pointer while trying to indicate their focus on a slide). For instance, users of the Wii console commonly point the Wii Remote at on-screen buttons, and Kinect-based applications often require the user to lift the hands so that their virtual counterparts reach the on-screen location of specific elements. Often, when aiming is not intended as a game challenge, on-screen elements are automatically highlighted to assist user selection, either by being the closest to the pointed location or by estimation of intent [12]. An approach that bears connections with the basic idea of this paper is one of using coded gestures to identify screen locations, rather than pointing. In effect, it is as if "pointing" would be split into two layers: the abstract gestural command ("pointing at x,y", "pointing at button 1", etc.), and the actual gesture (e.g., a specific elliptical motion [13] or rhythmic pattern [14] while on a specific screen).

As a final note, the precision shortcomings may also be part of system itself, rather than found at the user end, imposing this layered complexity upon the user. An example of this is the multitude of ways in which holding a Wii Remote differently while swinging it will produce spinning effects on a tennis ball in the game

Wii Sports [15], which bear little or no resemblance to the actual methods of imparting spin while using a real tennis racket. In fact, in this game a player can even hit the ball with particular spins by swinging the remote away from it, which is quite counter-intuitive.

2.2. Mimicking unnatural motion

While mimicking itself is complex, as seen above, quite often the actual movements cannot be mimicked effectively at all. An obvious case is that of flying, a relatively common action within virtual 3D environments. The typical approach is to employ trigger gestures to initiate or control flying, as demonstrated by Phan's sample control of Second Life avatars using Kinect, where lifting one's arms initiates flying [16]. Similar issues are present when a system requires swimming gestures, and in general for all instances when the virtual action on screen cannot be physically executed.

A complementary idea is found in the realm of arts and augmented dance expression, where a dance may intend for his/her virtual representation to perform unnatural body contortions or metamorphoses, not simply as triggered behaviors (which would be more akin to a button selection process as mentioned above), but as organic extensions to a dancer's moves and flow [17].

In both cases, the need for a distinction between gestures (executed by the users) and gestural commands (digital concepts perceived by the systems) is clear.

2.3. Non-kinestesic concepts

A final and more challenging category of issues is when there is no clear physical motion to express a command. I am referring to concepts for which mimicry is focused on concepts, analogies, and ideas – not on gesture imitation. For instance, what if one desires to pause a game? Standing still for a while may not be viable (a player could stand still to hide in a game, or simply be pondering the course of action or watching the virtual scene; a business user could be reading a long text or watching a presentation; etc.). What if one starts running on the same spot to order a virtual avatar to run, but wishes the avatar to proceed, rather than be forced to run as much as the virtual avatar? How does one issue a command for "keep going"?
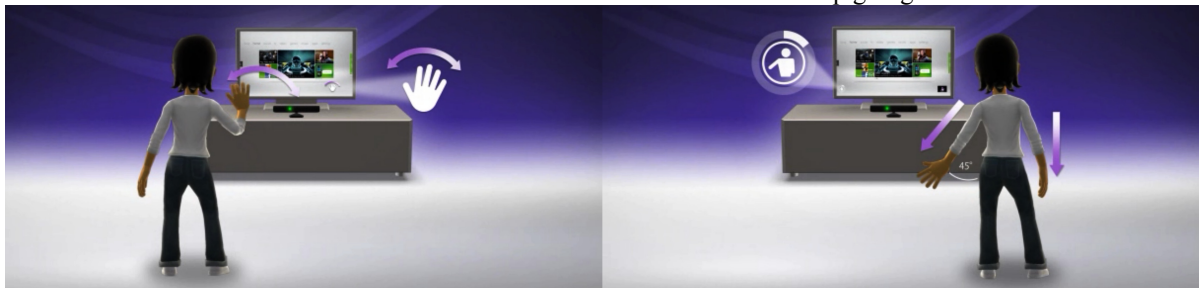

Fig. 4. Kinect non-imitation gestures: a) Wave to Kinect; b) Guide gesture
Video stills from http://support.xbox.com/en-US/xbox-360/kinect/body-controller

The typical response to these issues is to leverage the metaphorical richness of human gestures [23]. This is usually achieved by defining the gestural equivalent of escape sequences: arbitrary gestures which are meant to represent these non-physical concepts, rather than mimic an intended action. The set of standard gestures used by Microsoft Kinect provides a few examples of this (Fig. 4). For instance, to ask the Kinect sensor to start recognizing the user's presence, one should wave to the Kinect, as if saying "hello" to a virtual person; to issue a "pause" command, one needs to perform the "Guide" gesture, as a tour guide diverting visitors or a traffic

office telling a driver to steer to the side of the road. In both cases, the gesture is being directed at the Kinect-using system itself, not to the game or application being controlled. And thus their meaning is a mere convention, which needs to be taught and learned in order to be effective.

Another example is found in a video by Thai Phan, and engineer at MxR Lab, demonstrating the use of the Kinect to control a Second Life avatar [16]: in it, Phan show his solution for issuing zooming and panning commands for the viewport (rather than for his avatar): by pretending to hold a "window pane" in front of him, he approaches that imaginary window pane to the screen to zoom in, and moves it aside for panning (Fig. 6). Once again, this is an arbitrary gesture that needs to be taught and learned. And in the WiSee prototype, a set of *ad hoc* gestures is used as a marker to indicate that subsequent gestures are meant to be commands [9].
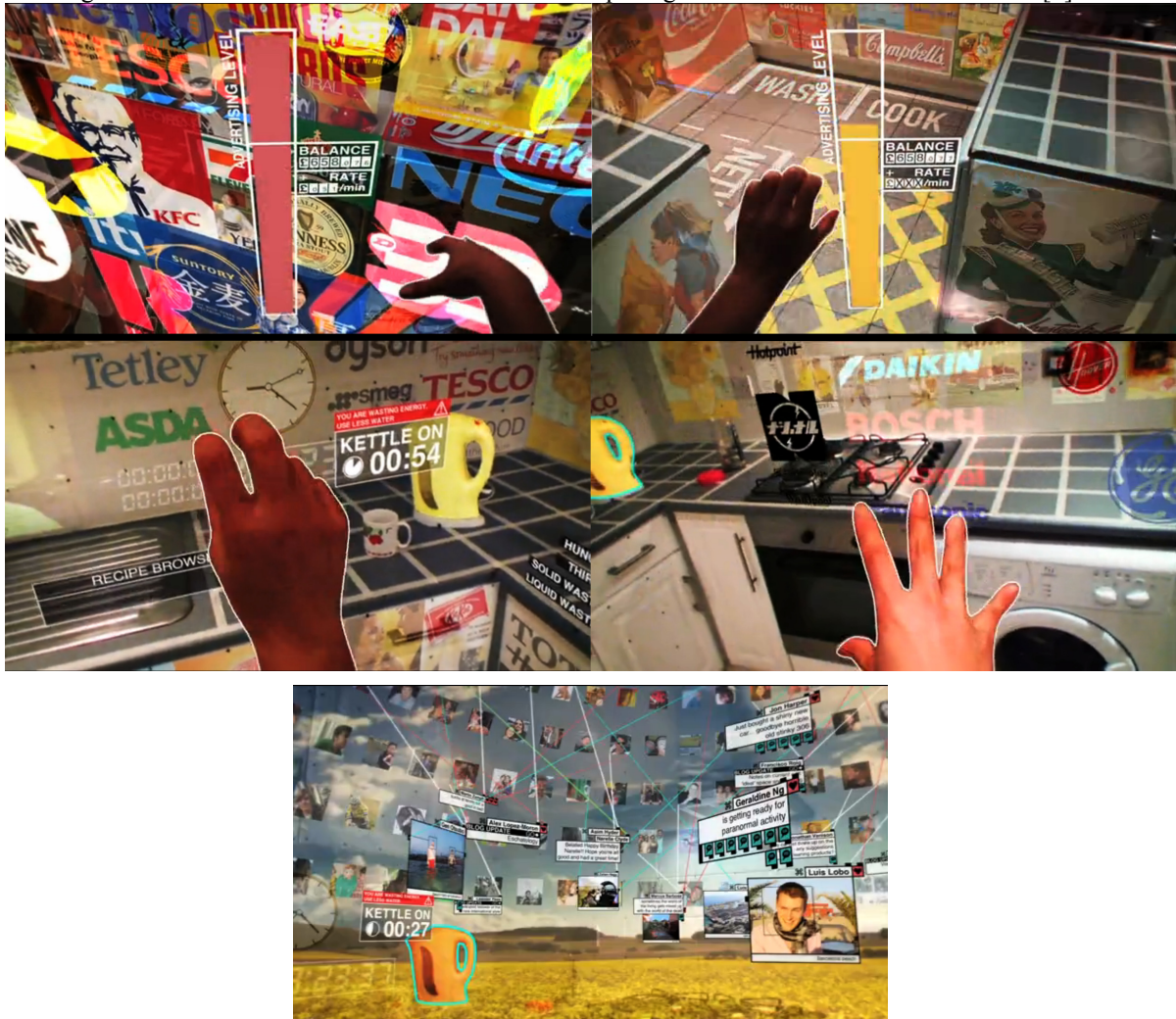

Fig. 5. Arbitrary gestural commands in the video "Augmented (hyper)Reality: Domestic Robocop". a) accessing advertising control slider; b) setting slider value; c) pulling up a recipe; d) sending away the current overlay; e) virtual meadow with social media posts.
Video stills from http://www.keiichimatsuda.com/augmented.php

   Possibly the extreme example of this is found in Keiichi Matsuda's architectural studies proposals called Domesti/City, for interaction with augmented city and domestic spaces [18]. In its companion videos, several non-kinestesic commands are issued via arbitrary gestures. Fig. 5 provides a few examples: in picture a) the gesture resembles the turning of a knob. It elicits the appearance of a slider control, labeled "Advertising level". This slider is then operated with the left hand, as shown in picture b). Not by pointing and dragging, but rather by a quick pull back of the wrist, as if pulling a large mechanical horizontal lever – this diminishes the advertising level, enabling the user to see the surroundings: a kitchen. The video then proceeds to show the user preparing a cup of tea, which involves getting a recipe. In picture c), to invoke the recipe, the user makes the gesture of pulling a filing card out of an imaginary filing drawer. Later in the video, the user decides to access a virtual meadow overlaid with social media posts. To access it, the gesture in picture d) is made: a quick opening of an outstretched hand, which triggers an animation of the current surrounding "flying away" and replaces them with the virtual meadow (shown in picture e). While one can quickly provide metaphoric explanations for these gestures, they are entirely arbitrary: there is nothing in the interface or the user's intent that would directly and intuitively lead him/her to perform these specific gestures. They need to be taught and learned. Possibly more eloquent as a demonstration of the challenge of gestures as commands for non-kinestesic concepts is the fact that in this video no gestures are shown during the time when the user is watching the social media posts in the virtual meadow.

Fig. 6. Making the gesture of holding a window pane in front of the screen for zooming and panning. Video stills from [16].

3. An inspiration from science fiction: the shamanic interface

   The need for users to learn specific arbitrary gestures associated with non-kinestesic concepts in order to command augmented reality or motion-based systems holds a potential learning challenge. Even if users can define their own sets of gestures, the potential for escalating complexity is high, as more and more gestures are created to achieve more and more commands.
   A possible approach to tackling this learning challenge and taming this escalating complexity was presented in a science fiction novel by Daniel Suarez in 2010, "Freedom™" (which is the second half of the story initiated in "Daemon", a previous volume). In it, the characters devised by Suarez use gestures to control an augmented reality interface [19]. In his words (*ibid.*):

> *«The shamanic interface is the mechanism for interacting with the darknet. It's called the shamanic interface because it was designed to be comprehensible to all people on earth, regardless of technological level or cultural background." She made a series of precise flourishes with her hands, leaving behind glowing lines* (…) *that formed an intricate pattern. As she finished, an unearthly, angelic voice sounded in the room* (…) *"It was a hypersonic sound, Sergeant. Linked to a macro that I created based on somatic gestures. But my point is that it looks like magic. Even the most remote tribes in Papua New Guinea understand the concept of magic—and that certain rituals must be observed to invoke it. They believe in a spirit world where ancestors and supernatural beings watch over them. The shamanic interface simply connects high technology to that belief system, granting 'powers' and equipment as a reward for useful, organized activity.»*

Suarez's idea seems to point towards a utopian single set of gestures for all – a stance which I don't subscribe, not least due to motion handicaps of many, but also due to culture's critical influence in determining the pattern and type of bodily movements – as demonstrated as early as 1941 by Efron for conversations [22]. Rather, my proposal in this paper is based on the idea that while such a single set of gestures for all is not viable, the kernel idea of linking gestures to the cultural beliefs of users may hold the key to rendering gestural commands of non-kinestesic actions easily learnable and readily remembered. Hence, this paper's proposal might be better named the "anti-shamanic interface."

4. Sample culturally-tied gestures for virtual interactions

Indeed all cultures include a rich variety of gestures that do not directly mimic actions, some directly devised for communication, some employed almost intuitively in support or for emphasis of oral speech, some purely symbolic, known as "emblems" [25]: anthropology as a discipline has developed both methods and field work to identify sets of such gestures (e.g., [24]). In order to explore the idea of exploiting this gestural richness for commanding computer systems, in this section I present two examples of gestures found in cultural traditions around the world and explain how they might be employed for interaction with virtual environments. These are mere examples to clarify the concepts herein, not actual proposals – those would require the identification of the set of commands required for interacting with a specific system.

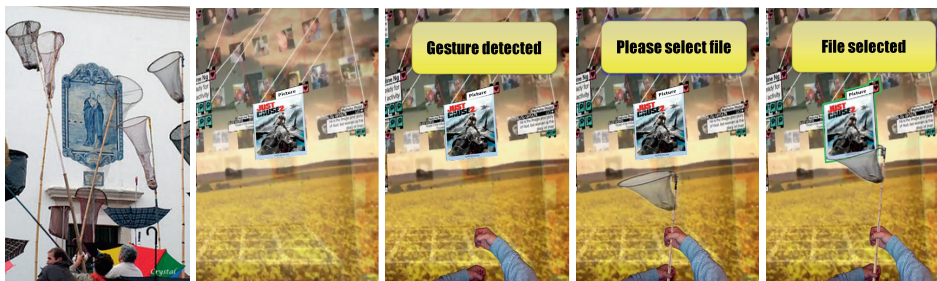### 4.1. Aveiro, Portugal: Saint Gonçalinho celebration

Fig. 7. a) Collecting cakes thrown from the chapel; b-e) Using the "pole holding" gesture to collect a file from the augmented reality space.

It is an yearly annual tradition at the Saint Gonçalinho chapel in Aveiro to offer cakes called "cavacas" by throwing them from the top of the chapel to the public below – the so-called "cake-hunters", which will employ any means to catch them, including umbrellas or bare hands, but also often using fishing nets on poles (Fig.

7a)[†]. One might thus use, for these users, the "pole holding" gesture to trigger the rendering of a virtual fishing net on a pole, which would then be used to "collect" an on-screen element, as demonstrated in Fig. 7b-e.

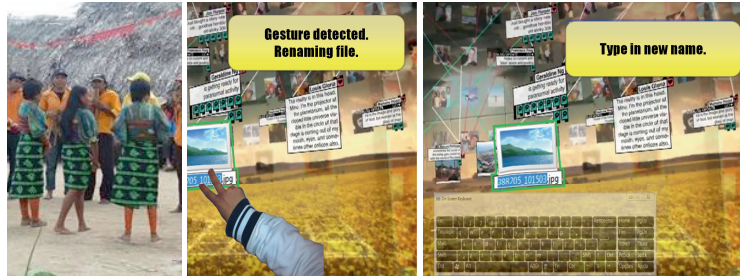*4.2. Cuna people, Panama: inna suid, rite of cutting a girl's hair*

Fig. 8. a) Girls during the hair cutting ceremony (notice central girl with long hair)[‡]; b) Using a "cutting" gesture to initiate renaming; c) Typing in the new name.

Among the Cuna people of Panama, several rites mark a girl's growth. Among them, the *inna suid* lasts between one and four days and is a celebration to cut a girl's hair (Fig. 8a). During this celebration, the girl receives her official Cuna name, which she'll use for the rest of her life [20] (p. 166). One might thus use, for these users, a "hair cutting" gesture over a file name in order to initiate a renaming process (Fig. 8b-c).

5. Final thoughts on possible impacts on accessibility and user training ease

To summarize, there are two components to the idea proposed herewith. The first is that since gestures for command of systems can go beyond mimicry and encompass non-kinestesic concepts, by embracing the cultural richness and depth of human cultures it is expectable that this approach may be found to be easier to learn and remember, by providing users with more direct links between existing societal meanings and new meanings in system interaction. Given its non-reliance on textual cues, this approach may also provide an opportunity for more advanced interaction on the part of illiterate users, including young children. Further, it may be found to be more adequate for widespread use of augmented reality interfaces, since it provides a pathway for issuing commands without voice intervention or keyboard gadgets, thus potentially becoming less conspicuous and causing less interference with other people in the vicinity.

The second component is that since cultures are immensely diverse, and users themselves have different motion restrictions, either due to impairments or context (i.e., gestures meant to be performed in confined spaces vs. gesture for wide spaces), gestures need to be decoupled from the systems that aim to use them for control. Thus, system architectures should have a set of abstract "gestural commands" and a software interface for receiving them, but decoupled from the actual identification of somatic gestures. Gesture-identification modules should use spatial data to identify a gesture and convert it into the actual "gestural command" the system requires. This would have to be a continual process, since gestures will often be dynamic, providing parameters during their execution, not simply identifying a command. But this decoupling, this creation of "cultural settings" for gestures alongside the now common "regional settings" found in most computing

---

[†] More info and sample video at http://www.guiadacidade.pt/pt/art/festa-de-sao-goncalinho-19648-01
[‡] Photo from http://blog.mailasail.com/kahia/163/image/jpgYgErGQj_SY.jpg

systems, would hold the potential foreseen here. The decoupling would also support other approaches, beyond the one proposed in this paper, such as personal definition of custom gestures.

Acknowledgements

References

[1] Microsoft (no date). *Kinect for Xbox 360*. http://www.xbox.com/Kinect (2013-05-23).
[2] Microsoft (no date). *Kinect Adventures!*. http://www.xbox.com/KinectAdventures (2013-05-23).
[3] Champy, A. Elements of Motion: 3D Sensors in Intuitive Game Design. *Analog Dialogue* 2007; 41-04;1–3
[4] Nintendo (no date). *Wii Official Site – What is Wii?*. http://www.nintendo.com/wii/what-is-wii (2013-05-23).
[5] Sony (no date). *PlayStation®Move Motion Controller - PlayStation®3 Move Info, Games & Updates*. http://us.playstation.com/ps3/playstation-move/ (2013-05-23).
[6] Lukowicz P, Amft O, Roggen D, Cheng J, On-Body Sensing: From Gesture-Based Input to Activity-Driven Interaction, *Computer* 2010, 43-10; 92-96.
[7] Leap Motion (2013), *Leap Motion*, https://www.leapmotion.com/product (2013-05-23).
[8] Thalmic Labs (2013). *MYO – The Gesture Control Armband*. https://www.thalmic.com/myo/ (2013-05-23).
[9] Pu Q, Gupta S, Gollakota S, Patel S, Whole-Home Gesture Recognition Using Wireless Signals. To appear, at *The 19th Annual International Conference on Mobile Computing and Networking (Mobicom'13)*. Referenced from the on-line version at http://wisee.cs.washington.edu/wisee_paper.pdf (June 15th, 2013).
[10] Gerling K, Livingston I, Nacke L, Mandryk R. Full-body motion-based game interaction for older adults. In *CHI'12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: ACM; 2012, pp. 1873-1882.
[11] Vatavu R, Nomadic gestures: A technique for reusing gesture commands for frequent ambient interactions, *Journal of Ambient Intelligence and Smart Environments*, 4-2, 79-93.
[12] Bieg HJ, Chuang LL, Fleming RW, Reiterer H, Bülthoff HH. Eye and pointer coordination in search and selection tasks. In *ETRA '10 Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, New York: ACM; 2010, pp. 89-92.
[13] Fekete JD, Elmqvist N, Guiard Y. Motion-pointing: target selection using elliptical motions. In *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: ACM; 2009, pp. 289-298.
[14] Ghomi E, Faure G, Huot S, Chapuis O, Beaudouin-Lafon M. Using rhythmic patterns as an input method. In *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York: ACM, 2012, pp. 1253-1262.
[15] Wii Blog (2007), *Wii Sports - Tennis Shots Explained*, http://wiiinfo.blogspot.pt/2007/01/wii-sports-tennis-shots-explained.html (2013-05-23).
[16] Phan T (2011). *Using Kinect and OpenNI to Embody an Avatar in Second Life* [YouTube video], http://www.youtube.com/watch?v=ehTvtkybubM (2013-05-23).
[17] Valverde I (2012). *Sensesplaces – Mixed Reality Participatory Performance Environment*, http://sensesplaces.wordpress.com/ (2013-05-23).
[18] Matsuda K *Domesti/City – The dislocated home in augmented space*. Master thesis, Bartlett School, University College London.
[19] Suarez D, *Freedom™*, Boston, MA, USA: Dutton; 2010.
[20] Cheville L, Cheville R, *Festivals and Dances of Panama*, Self Published, 1977. http://books.google.pt/books?id=u3oNAAAAYAAJ&dq=editions%3ALCCN81176587
[21] Rehm M, Bee N, André E. Wave like an egyptian -- accelerometer based gesture recognition for culture specific interactions. In *BCS-HCI '08 Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 1*, Swinton, UK: British Computer Society; 2008, 13-22

[22] Efron D, *Gesture and Environment: A Tentative Study of Some of the Spatio-temporal and Linguistic Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living Under Similar as Well as Different Environmental Conditions*, Morningside Heights, NY: King's Crown Press; 1941.

[23] Cienki A, Müller C. Metaphor, gesture, and thought. In *The Cambridge Handbook of Metaphor and Thought*, Cambridge, UK: Cambridge University Press; 2008, ch. 27.

[24] Kendon, A (1992). Some Recent Work from Italy on Quotable Gestures (Emblems), *Journal of Linguistic Anthropology*, 2(1), 92-108.

[25] Johnson H, Ekman P, Friesen W (1975). Communicative Body Movements: American Emblems, *Semiotica*, 15(4), 335-353.