Translating Graphical User Interfaces: challenges for the design and standardization of mid-air interfaces

Tobias Mulling[1], Derek Covill[2], and Lyn Pemberton[3]

University of Brighton, Brighton, United Kingdom
[1] t.mulling@brighton.ac.uk
[2] d.covill@brighton.ac.uk
[3] lyn.pemberton@brighton.ac.uk

Abstract. This study presents a theoretical contribution on mid-air interface design. With the technological development of devices able to interpret gestures made by hands and arms as Kinect and Leap Motion, designers and developers might think beyond the traditional mouse and keyboard input. However, these interfaces that supposedly aim to be more natural and intuitive have found barriers to its acceptance by users and researchers, partly due to a lack of understanding of the variables involved in the project, as well as the lack of existing standards. This article aims to bring insight into the translation and understanding of these interfaces as language, establishing a communication process based on the following principles: (1) an understanding of Mid-air interfaces and its concept, (2) an analysis of the different ways to interact, (3) an understanding of native and intuitiveness of gestures and (4) generating insights for mid-air interface standards.

Keywords: mid-air interface, design patterns, gesture based interaction, freehand interaction.

Introduction

In 2002, the film Minority Report, starring Tom Cruise brings out a dynamic, innovative interface where using only hands, without touching the screen, one can organize information, manipulate different types of forms and data, in an intuitive and natural way. Despite being an interface present only in Sci-Fi, it represented the future of interfaces and interactions, a mid-air interface.

More than a decade has passed, and the present has revealed to be a little different than imagined by the film producers. The technology to simulate some of the predicted interactions is finally available to designers and developers, but the popularity of mid-air interfaces still does not occur on a large scale. Designers and developers have not yet met the standards for design of this type of interface, in addition, the actual application of the technology in specific niches is still unknown. This paper aims to present a theoretical contribution, in a reflective way, about paths taken in mid-air interface design, trying to show some possible mistakes and successes in the process of translation of graphical interfaces, seeking for a syntax for gestural language.

Mid-air interfaces

The primary instrument for human beings to communicate is the use of language. This may be spoken, written, and should follow a standard structure and convention (Oxford 2015) to enable the establishment of a communication process. Among these languages, one of the most typical ways to establish communication among humans is non-verbally, that is, by using gestures. From birth, humans begin to use us gestures, sometimes involuntary, but over time these gestures allow for an intention to take place, and they end up establishing communication when needed. For example, moving hands with the intention of waving good-bye, closing the hand and lifting up a thumb to communicate agreement to something (like gesture), moving an arm from one side to the other indicating where to put certain object, and even gestures combined with speech are examples of gestures made in the day-to-day lives. Based on that, it is plausible to say that gestures may also contain information, thus indicating that the use of gestures is a (kind of)

language. In this perspective, Billinghurst & Buxton(2014) define a gesture as "the motion of the body which contains information". Saffer(2009) explains the concept of gesture attributed to any physical movement detected by digital sensors, which generate a response without the aid of traditional mechanisms, such as a mouse and a keyboard. This definition reflects what happens in the application of gestures in human-computer interaction (HCI), which has in its concept the importance of detecting a gesture. The detection of a movement is associated with the use of gestures as a language, as a way of communicating with a system. As part of the human-computer interaction, the interaction design defines the structure and behaviour of interactive systems (IxDA 2015). Thus, to design a system capable of interpreting Sign language system it is necessary to understand the relations (interactions) that occur between the user of the system and the product itself, and that is mediated by an interface. Johnson reaffirms the importance and function of an interface, which “acts as a kind of translator, mediating between the two parties, making one sensitive to the other.”(Johnson, 1997, p18)

Generating the establishment meanings and by consequence the translation of an interface is understood by Saffer (2010, p. 4) as the "art of facilitating interactions between humans through products and services." While the gestural interaction from smartphones and tablets is already quite popular through touchscreen displays, this "language" is still maturing over the use of the movement of hands and arms (or even the whole body) to control different interfaces. This type of interface, known as mid-air interfaces have been used in games such as Microsoft Kinect (released in 2010) and recently by other devices such as Leap Motion (released in 2014) and MYO (released in 2015). However, its application in other contexts such as health software, education, entertainment is still in an embryonic stage. This fact may partly be associated with improved reading technology and interpreting gestures, as well as the beginning of the process of popularization of technology, which begins to allow for devices at a more affordable cost to users.

Research related to interpreting gestures has already been carried out for more than two decades (Buxton 2007). However, much of the effort and research regarding the interpretation of gestures have been related to technical aspects such as algorithms that could track the movements more accurately, sensor optimization, etc. So the challenge is how to design graphical interfaces (visual elements shown on the display) so that they are consistent with the paradigm of gestural interaction, intuitive and that provide a satisfactory user experience. Understanding gestural-interaction as a language, especially in a communication context, permeates the understanding of how potential users use gestural interaction, as well as the better way to translate this language to visual elements in the graphical interfaces, therefore allowing interaction.

Regarding the translation of these interfaces in addition to the lack of standards, there are few consistent frameworks for the design and evaluation of mid-air interfaces (Cabreira & Hwang 2015). Some authors associate the concept of mid-air interface to the perception of a natural user interface (NUI) and more intuitive, because the interaction is based on gestures (O’hara et al. 2013; Wigdor & Wixon 2011). It is suggested that this interaction paradigm may prove to lower barriers between users and interfaces (Grandhi et al. 2011). However, when using the same principles of construction of graphical interfaces that are used for different interaction paradigms, different mid-air developed interfaces have received criticism regarding its usability (Malizia & Bellucci 2012; Norman 2010). Therefore, translate and understand the mid-air interface design is presented as a process still under development, searching for answers and standards.

Different ways to interact

Throughout history, human-computer interaction has been established mainly by interfaces that support the process of interaction established between man and machine. Initially this interaction was established by interfaces of a category called "command-line interfaces", where the user would type commands and the software would respond by processing them. With the maturation of technology and enhancement of electronic graphical components there was the emergence of the first Graphical User Interfaces, or GUI, i.e., metaphors of user's everyday actions, making the interaction became friendlier. This strategy was responsible for the expansion of computers in homes and professional environments, popularising them on a global scale (Saffer 2010).

Coupled with the concept of having a user-friendly interface (GUI), the WIMP paradigm (short for "Windows, Icons, Menus and Pointers") still holds the majority of the graphical interfaces. This was initially used at the interfaces of the Xerox Alto computer (1973), Apple Macintosh (1984), and came to become popular mainly with computers that started to use the Windows operating system. This paradigm is mainly characterised by the use of metaphors: visual elements that have real object references. This strategy in the interface design was conceived in order to reduce the learning curve, linking the role of screen elements to their function in a non-digital environment (e.g. Trash to delete files).

The visual elements that make up the WIMP pattern are responsive to an interaction model that remains to this day from the use of the mouse as a navigational tool (Shneiderman & Plaisant 2010). Thus, the projected GUIs become responsive to point and click pattern, showing the user what is clickable or non-clickable from different forms of feedback, especially when the mouse is under an interaction element (known as hover).

With the release of iPhone in 2007, the touchscreen gestural interaction begins to become popular along with smartphones and tablets. Its main differences compared to interaction is the inclusion of gestures (swipe, pinch, etc.), allowing tasks performed in interfaces once mediated by mouse now to be rethought to a new model where the interaction could occur with 1, 2, up to 5 fingers, even allowing the combination of two or more fingers at the same time. Due to this new paradigm, some visual characteristics of the graphical interfaces become modified, partly because of the very evolution of graphical interfaces from websites, but also by standards set by own systems (Apple, Android and Windows Phone). Furthermore, the amount of information arranged in the users' screen becomes reduced due to the size of smartphones, also allowing a reduction of the cognitive load of the interface. This perspective in the creation of interaction design is also associated with the concept of the first mobile interfaces and system design. (Wroblewski 2011)

In gestural interaction using hands and arms (mid-air interface), the vocabulary used for gestures becomes enlarged, this is due to both the use of fingers as input data and the two hands, which are combined combination in various gestures (open, close, move the hand, for example). Thus, the variables involved in the design of an interface are also extended because there is a need to understand how this paradigm influences the design of different visual elements shown on the interface. A major difference between the input point and click (desktop computers) and mid-air interfaces is associated with the fact that, in most cases, devices that interpret gestures use a camera. Thus, this means that any user's movement in the device detection area is regarded as an information, allowing involuntary gestures to be interpreted as commands in an interface (a phenomenon known as mic live) (Chaudhary et al. 2011). This feature also directly affects the interface in relation to visual elements; certain navigational gestures are likely to be triggered by the user movement of hand (even if unintentionally). In order to getting the hang of this feature, try browsing a website without clicking, trying to figure out different ways to click / activate a button. Furthermore it is important to reflect on the choice of application gestures set to avoid false positives (Wigdor & Wixon, 2011) or ambiguous gestures, or gestures that get closer as they are being executed and that may be confused by the system with another gesture. In the picture below you can see the gestures of an application and the transitions between them, showing possible points of ambiguity and / or false positives.

Another difference in mid-air interfaces, especially in relation to the touchscreen / mouse is the lack of haptic feedback. The touchscreen interfaces, although they have a paradigm of gestural interaction, frequently use the gesture "tap" as it follows the standard point and click of the mouse. In mid-air interfaces this standard should be rethought, precisely because it is not possible to click buttons or simply tap surfaces (Ren & O'Neill 2012). Therefore, it is necessary to reflect on the navigation strategy used, so that the graphical interface presented to the end user represent a coherent mix of visual elements that may trigger the correct gestures designed for each specific command. Also, considering which metaphors / patterns of graphical desktop paradigm or touchscreen interfaces should be kept / modified, or redesigned in order to optimize the user experience in mid-air interfaces.

Naturalness and Intuitiveness of gestures

Conceptually, this idea of mid-air interfaces (NUI or touchless) seems to be promising, as it can optimize satisfaction for the use of a system, as it allows the user to use his/her own language, gestures, movements (appropriated during his

life) to interact with an interface. However, although some gestures can are commonly used by most of the population, this fact does not guarantee that the interface of a system might be "natural", simply because it utilizes gestures. At some point the user should probably learn this gesture to later pass to use it for granted. Herein lies one of the main discussions of the term "natural interface"; the so-called interfaces try to simulate the interaction of day-to-day lives or simply have a higher degree of freedom and expression when compared to a mouse or keyboard (Malizia & Bellucci, 2012). Furthermore, a gesture may have different cultural aspects, i.e. be interpreted in different ways by different people. Under the interaction design, Norman (2010) also suggests that the "natural user interfaces are not natural," because they do not follow some basic rules of human-computer interaction, especially related to feedback and navigation.

Because of criticism against the concept of mid-air interfaces and its application, Malizia & Bellucci (2012) state that the initial idea of mid-air interfaces would be to identify possible actions and pre-existing gestures of people, identified from people's manipulation with different objects and tasks, for, then transport them to the digital environment. However, part of these interactions are associated with a context which sometimes ends up getting lost in the digital environment. The naturalness of these interactions is something that is taken as purely a problem of representation – ensuring that they are correctly represented in the interaction mechanism itself. In this sense, natural interactions are something "detached from the social context in which they might be deployed; they are not constituted by the context, but brought to it." (Malizia & Bellucci 2012)

Regarding the intuitiveness of gestures, the literature indicates that dynamic gestures - gestures that have movement from an initial position to a final position - have preference with users. The use of pantomimic gestures is also considered more intuitive and easy for users (Grandhi et al., 2011). As an example, when simulating an activity to cut an object, users tend to move hands up and down, with the open palm, simulating the cutting of a real object. However, part of this research does not list the execution of the gesture with the display of the GUI, but represent a reaction of users to an image or given scenario. Another possible problem is that pantomimic gestures and symbolic gestures are often associated with culture, and may thus differ from region, customs, etc.

As an alternative to the process of translation and design of graphical interfaces for the mid-air type interactions, Ardito et al. (2014) suggest the use of manipulative gestures, i.e., from direct manipulation of elements in the interface provide gestural navigation. According to the authors, rather than causing users to learn and / or remember a vocabulary of gestures so often artificial, it would be more intuitive if the user could directly manipulate the interface elements from the recognition or discovery of a designed function. The native gesture might be improved through the use of an object or interface element that refers to a metaphor of an object or daily action of the day-to-day users.

Looking for Patterns

As already mentioned, with the emergence of the first graphical user interfaces (GUI) and standard WIMP interfaces have become more user friendly, and began to popularize. One reason for this popularity is the use of interface standards. They describe design solutions to be applied in specific contexts (Pauwels et al. 2010). As an example, the organization of content in tabs, or the carousel menu type, have different ways of organizing and allow navigation in content. The design patterns express a language, and that happens to be associated with the interaction paradigm. The interfaces using the mouse as interaction (desktops and laptops) already have a large library of patterns established. With the development of touch screen interfaces, there was a translation of standards for this interaction paradigm. As a result, some patterns have changed and new standards go to be created and adopted by designers and developers.

The vast majority of these patterns follow the model WYSIWYG (What You See Is What You Get). However, with the popularization of interfaces for smartphones, vocabulary standards begin to contemplate gestures and, as mentioned, different from the standard point and click, one can use many gestures to the interaction. This fact has gradually allowed the insertion of standards that are not displayed in the user interface, but activated from the use of gestures.

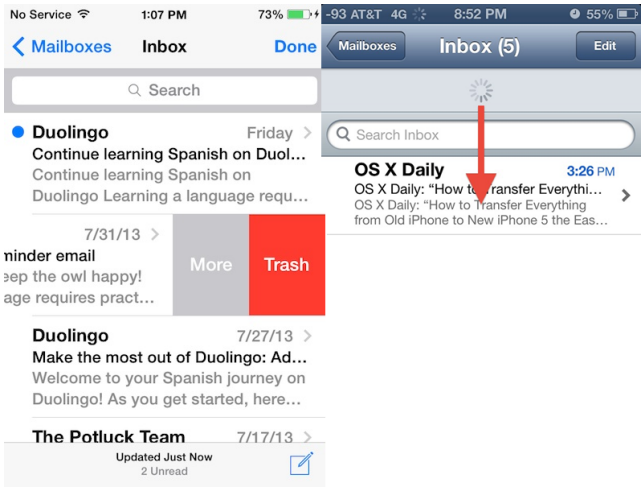

Figure 1. Swipe gesture to right (reveal more options) and Swipe down (refresh mail)

Examples in Figure 1 show two interface standards that are viewed only from the gesture. In the first example (left image), the options of "more" and "trash" only appear after the gesture swipe left. The same applies to the second example with the “refresh” action (right image); after you perform the gesture swipe down, the interface provides the feedback, so checking for new emails. The mid-air interaction still goes through a maturation and discovery process. With the work of designers and developers , as well as testing with users, it is expected that interface patterns will be designed and validated . Thus, based on the observation of the mid-air paradigm and reflection on interfaces, the following table presents insights that can assist designers in the development of these applications taking into account differences in relation to the interaction with the cursor paradigm.

Graphical User Interface (Desktop Paradigm)	**Mid-air Interfaces (Gestural Paradigm)**
Control is given by mouse or keyboard	Direct Manipulation using hands
More accuracy, less immersion	Less accuracy, more immersion
Makes use of mental pattern already used at previous softwares related to computing experience	Makes use of users assumptions, discovery and logic conclusions
Emotionally related to work	Emotionally related to entertainment
Ideal for task productivity and efficiency	Ideal for social and collaborative tasks
Interface is visible and graphic	Interface is physic and can be either visible or invisible
Interface depends on targets (buttons, menus, etc)	Interface focuses on gesture, not necessarily dependent on an specific position on the visible screen
Long-term duration	Fast-execution interaction and short-time duration (except for games)
Fatigue is felt after long-term interaction	Fatigue attributed to gestures which imitate the movement of cursor (gorilla arm)
Browsing with different hierarchy patterns	Browsing with few hierarchy patterns
Naturality in interaction is at the cursor	Naturality in interaction is at the relationship between gesture and interface

Table 1. Insights and differentiation between desktop and mid-air interfaces

Conclusion: translating mid-air interfaces?

With the development of gestures interpreting-able devices, it is up to designers and developers to understand in more depth the variables that make up this type of project. One possible way to translate this type of interface may become

possible through the observation of how the process of creating gestural interfaces for smartphones took place, and from the touchscreen interfaces seek clues to the definition of standards for mid-air interfaces. This does not mean copying the visual elements and interaction process, but to understand and build a visual and gestural language for the design of mid-air type applications. This paper aims to demonstrate some important concepts related to this paradigm of interaction, seeking to generate insights for translating and mid-air interface design.

Additional Information

Acknowledgements. This work was undertaken with support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, grant Doutorado Pleno no Exterior–Proc. BEX BEX 13430/13-0.

References

Ardito, C., Costabile, M.F. & Jetter, H., 2014. Gestures that people can understand and use. Journal of Visual Languages & Computing, 25(5), pp.572–576.

Billinghurst, M. & Buxton, B., 2014. Gesture Based Interaction. In Haptic input. pp. 1–35.

Buxton, B., 2007. Multi-Touch Systems that I Have Known and Loved. Available at: http://tinyurl.com/2vl99n [Accessed November 24, 2014].

Cabreira, A.T. & Hwang, F., 2015. An analysis of mid-air gestures used across three platforms. In Proceedings of the 2015 British HCI Conference on - British HCI '15. New York, New York, USA: ACM Press, pp. 257–258.

Chaudhary, A. et al., 2011. Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey. International Journal of Computer Science & Engineering Survey, 2(1), pp.122–133.

Grandhi, S.S.A., Joue, G. & Mittelberg, I., 2011. Understanding naturalness and intuitiveness in gesture production. In Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11. New York, New York, USA: ACM Press, p. 821.

IxDA, 2015. Definition of IxD. Available at: http://tinyurl.com/pm9y4sx [Accessed August 11, 2015].

Johnson, S., 1997. Interface Culture: How New Technology Transforms the Way We Create and Communicate, Basic Books.

Malizia, A. & Bellucci, A., 2012. The artificiality of natural user interfaces. Communications of the ACM, 55(3), p.36.

Moggridge, B., 2006. Designing Interactions, The MIT Press.

Nacenta, M. a et al., 2013. Memorability of pre-designed and user-defined gesture sets. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.1099–1108.

Norman, D., 2010. Natural user interfaces are not natural. interactions, pp.6–10.

O'hara, K., Harper, R. & Mentis, H., 2013. On the naturalness of touchless: putting the "interaction" back into NUI. ACM Transactions on …, 1(212), pp.1–27.

Oxford, 2015. Oxford Dictionnaires. Oxford Dictionnaires. http://www.oxforddictionaries.com/ [Accessed August 19, 2015].

Pauwels, S.L. et al., 2010. Building an interaction design pattern language: A case study. Computers in Human Behavior, 26(3), pp.452–463.

Ren, G. & O'Neill, E., 2012. 3D Marking menu selection with freehand gestures. IEEE Symposium on 3D User Interfaces 2012, 3DUI 2012 - Proceedings, pp.61–68.

Saffer, D., 2010. Designing for Interaction: Creating Innovative Applications and Devices 2nd ed.,

Saffer, D., 2009. Designing gestural interfaces, Beijing: O'Reilly.

Shneiderman, B. & Plaisant, C., 2010. Designing the user interface: strategies for effective human-computer interaction, 5th ed., International ed., Upper Saddle River, [N.J.]: Addison-Wesley.

Wigdor, D. & Wixon, D., 2011. Brave NUI world: designing natural user interfaces for touch and gesture 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Wroblewski, L., 2011. Mobile First, Available at: http://www.lukew.com/ff/entry.asp?933.